

**OXFORD BROOKES UNIVERSITY
SCHOOL OF EDUCATION**

**Thesis towards Doctor of Education (PR3)
Degree awarded by Oxford Brookes University
Submitted April 2003**

**Evaluating Process and Outcome in the
Education of General Practitioners**

DR. EDWARD BASIL PEILE

**The Following Table and Appendices
have been omitted on request of the
University –**

Tables 3.2 & 3.3 pg. 34

Table 3.7 pg. 37

Table 5.6 pg. 92

Appendices 9-11 pg. 176-190

Appendix 13 pg. 203

ABSTRACT

I set out to develop methodologies linking the educational processes adopted by General Practitioner trainers to outcomes, in terms of quality performance by their learners in later life as doctors.

Evidence about educational process and about quality of practice must be collected and analysed in a format that takes full account of the judgements to be made for formative and summative assessment. This work iterates between considerations of evidence and judgements

The first of three phases of research established a framework of categories and dimensions by which to describe educational behaviours of GP Trainers. This involved interviewing trained practitioners to find what had been of lasting value from training. The categories deriving from a Grounded Theoretical approach have proved useful in practice and have been incorporated into training assessments.

Second Phase Research involved refining data collection methods for assessing prevalent educational behaviours in training practices. The process by which judgements are made about training was analysed and developed in the light of research findings, which support a trend towards self-assessment by trainers. The complexity of evidence collection is such that even experienced visiting teams struggled to construct meaningful aggregations across several categories in the course of a brief visit. Their limited data best serves to validate the self-assessments of trainers, carried out over an extended period of training, and involving potentially beneficial reflection-in-action and reflection-on-action.

Finally, 31 General Practitioners engaged in a pilot study of Insight 360® assessments of quality practice. Their self-assessments were compared against 331 patient assessments and 237 colleague perceptions. Literature review and preliminary experiments led to the conclusion that Multilevel Modelling (MLM) techniques are best suited to such data analysis. Even with small numbers, valid findings emerged around gender influences on self-perception, reinforcing the conclusion that MLM is needed if we are to relate complex data around quality of practice to the level of prior educational experience.

Using the framework developed in this project, trainers can now be encouraged to examine their prevalent educational behaviours and record the evidence for formative and summative assessment. This work gives confidence that accumulated 360° assessments of practitioners may in future be analysed using MLM techniques to shed light on different quality outcomes of varying educational processes.

ACKNOWLEDGEMENTS

I would like to acknowledge my gratitude for the NHS (SE) Research and Development Training Fellowship. The award for research in medical education was, I believe, an acknowledgement of the contribution which effective research-grounded education has to make to patient care. This funding enabled me to undertake studies for the Doctorate of Education, at Oxford Brookes University.

My academic supervision for the Doctorate has been of the highest order. Dr Joe Harkin, Reader in Education at Westminster Institute of Education, Oxford Brookes University has been supervising me from the start. Professor Helen Bartlett moved from the School of Healthcare at Oxford Brookes University to Australia, part way through the project, and Dr Paul Yerrell, who succeeded her as Director OHCRA, kindly took over my supervision. It has been extremely helpful for me to have supervisors from Schools of Education and Health working in close collaboration.

Throughout this project I have received tireless support from my Head of Department at the Department of Primary Healthcare, University of Oxford. Dr Regina Conradt has been a great help to me, in particular with the Multi-level modelling and statistical analysis. Dr Vanessa Simonite, of Oxford Brookes University, has also been invaluable in providing statistical help.

I have been extremely lucky to have had tutorials with Dr Neil Johnson, who, at the outset of the project was Director of General Practice Training at Oxford-PGMDE. Despite two moves on promotion, Neil has managed to fit in regular tutorials, which, without exception, have proved stimulating and helpful.

Charlotte Schembri has done more than help me with secretarial support, and proofreading. She has tried to keep me sane through all the trials and tribulations of computer crashes, and supported my caffeine intake.

As for my family, and in particular my wife Linda, I am eternally grateful for the tolerance of my time out of family life whilst I have been engaged in the research, and for the whole-hearted support which has been essential to seeing the project through.

CONTENTS

Chapter 1	Introduction	
	The Thesis Structure and Theoretical Issues	
1.1	Structure of the thesis	7
1.1.1	Thesis Overview	8
1.1.2	Concept Chapters	8
1.1.3	Measurement chapters	8
1.1.4	Reflection chapter	9
1.2	What is so elusive about process and outcome?	9
1.3	Performance, Perceptions, and Assessments in General Practice Education	10
1.3.1	The Good Doctor	10
1.3.2	Performance as distinct from Competence	10
1.3.3	Performance	11
1.3.4	Competence	12
1.3.5	Perceptions in assessment	13
1.3.6	Appraisal and assessment defined	13
1.4	Actions, Evidence about them, and Judgements upon them	14
1.4.1	The actions which are the subject of report	14
1.4.2	Evidence of observers	15
1.4.2.1	Perception formation	15
1.4.2.2	Perception recall	16
1.4.2.3	Perception transmission	17
1.4.2.4	Aggregation of perceptions	18
1.4.3	Judgements	18
Chapter 2	The Context of General Practitioner Training	
2.1	GP Registrar training	19
2.2	General Practice Trainers: Accreditation and Development	20
2.2.1	Accrediting the Trainers nationally	20
2.2.2	Accrediting the Trainers locally: peer review visits	21
2.2.3	Accrediting the Trainers locally: Team-leaders' reports	22
2.2.4	Accrediting the Trainers locally: Trainer Selection Committee	22
2.2.5	Educational needs of GP Registrars – what is evidence?	23
2.2.6	Evidence about the impact of GP training on subsequent performance in Practice	23
2.3.1	Perceptions in Medicine and Medical Education: Perceptions of ex-trainees	25
2.3.2	Perceptions in Medicine and Medical Education: Patient perceptions	26
2.3.3	Perceptions in Medicine and Medical Education: Perception of colleagues	26
2.3.4	Perceptions in Medicine and Medical Education: Self-perception	28
2.3.4.1	“360 degree” appraisal	28
2.4	Assessments of clinical practice	28
2.4.1	Attending to under-performance	30
2.4.2	Appraisal and revalidation for all doctors	30
2.4.3	Assessments of educational practice	30
2.4.4	Evaluations of performance across the medical educator's roles	31

Evaluating Process and Outcome in the Education of General Practitioners

Chapter 3	Phase One Research	
3.1	Conceptual: Phase One Design	32
3.1.1	Design problems for categorising process	32
3.1.2	Possible approaches	33
3.1.3	Choice of Research Subjects	39
3.1.4	Choice of Interview Method	40
3.1.5	Choice of Methodology	40
3.2	Measurement: Phase One Method	40
3.2.1	Category 1: Training or Education	43
3.2.2	Category 2: Style Spectrum	43
3.2.3	Category 3: Space for Reflection	43
3.2.4	Category 4: Modelling Personal Development and Teams	44
3.2.5	Category 5: Learning Cycles	44
3.2.6	Category 6: Family practice in context	45
3.2.7	Category 7: Control and Direction	45
3.2.8	Category 8: Feedback	45
3.3	Reflection: Pulling the Phase One threads together	46
3.3.1	Relating the Phase One research findings to other General Practice Literature	46
3.4	Summary	47
Table 3.1	Advantages and disadvantages of importing and adapting a framework	33
Table 3.2	Buchler's Typology of Method	34
Table 3.3	Craft Orientated Approach to Competence in Education	34
Table 3.4	Matrix of Method against craft orientated approach to competence	35
Table 3.5	Dreyfus & Dreyfus levels of proficiency	36
Table 3.6	Benner's Domains of Nursing Practice	37
Table 3.7	Matching Learner Styles to Teacher Stages	37
Table 3.8	Advantages and disadvantages of designing a framework conceptually	38
Table 3.9	Advantages and disadvantages of starting afresh; grounding theory on research findings	39
Table 3.10	Categories and Dimensions of Educational Behaviours of lasting value to learners	42
Chapter 4	Phase Two Research	
4.1	Conceptual: Phase Two Design	48
4.1.1	Sequential stages of Phase Two	48
4.1.2	Limitations imposed by research context	48
4.1.3	Compromises in the research endeavour	49
4.1.4	Accrediting the Trainers locally: Personal experience of the researcher	49
4.2	Measurement: Phase IIa (Understanding how judgements about teaching are made on the basis of inspection reports)	51
4.2.1	Comparison with accreditation approval outcomes	52
4.3	Measurement: Phase IIb (Pilot)	54
4.3.1	Research participant Trainers	55
4.3.2	Research Visits and Methods of Assessment	56
4.3.3	Joint Pilot Visits	59
4.3.4	Solo Pilot Visits	59

Evaluating Process and Outcome in the Education of General Practitioners

4.3.5	Designing a Training Package	61
4.3.6	Incorporating the methods into an actual re-accreditation visit	61
4.3.7	Training the Trainers	61
4.4	Measurement Phase IIc (Roll-Out)	62
4.4.1	Routine Assessment of training behaviours for a trial period	62
4.4.2	Team-leader's study day	64
4.5	Reflection Phase IIa	65
4.5.1	How consistent are individual judgements	65
4.5.2	How similar are experienced assessors in their grading	65
4.5.3	Assessments and perceptions: discussion in committee – the 'shadow' effect on individual perceptions	65
4.5.4	Assessments and perceptions: discussion in committee – are all perceptions equally weighted?	67
4.5.5	Assessments and perceptions: discussion in committee - can we truly exclude prior knowledge from the 'shadow' effect?	67
4.5.6	How can we improve the fairness and reliability of committee discussion?	68
4.5.7	How can we improve the reliability of Team-leaders' reports?	69
4.5.8	Conclusions from researching Team-leader reports and the assessment process	69
4.6	Reflections Phase IIb and IIc	70
4.6.1	Standard Scenarios reconsidered	71
4.6.2	Context specificity issues	71
4.6.3	Conflicting interests of research and quality improvement	72
4.6.4	No proof that preferred behaviours are better	73
4.6.5	Assessing longitudinal process in cross-section	74
4.6.6	Ethical Considerations	75
4.7	Summary	76
Table 4.1	Options for research methods Phase Two	50
Table 4.2	Key to Committee Decision Outcomes	52
Table 4.3	Graded outcomes of committee decisions by prior (individual) ratings	54
Table 4.4	Different observers rating Team-leaders reports	53
Table 4.5	Examples of opportunities for assessing training behaviours on re-accreditation visits to training practices	57
Table 4.6	Team-leader's recording of dominant educational behaviours on Assessment Visits	64
Table 4.7	Comparing the two most experienced observers	65
Table 4.8	Concerns about 'Prior knowledge' affecting Committee decisions	68
Chapter 5	Phase Three Research	
5.1	Conceptual: Design Considerations Phase Three (Methods)	77
5.1.1	Attributes desirable in an assessment process	77
5.1.2	Measures of quality doctoring	77
5.1.2.1	Quality Marker Options for this research	79
5.1.3	Research participation has to be feasible in line with the constraints on GP time	80
5.1.4	The workload in General Practice	80
5.1.5	The Case for a Perception-based Research Instrument	82
5.1.6	Research Instrument 1: 360 degree assessment	82
5.1.7	The case for a portfolio of Competence	83

Evaluating Process and Outcome in the Education of General Practitioners

5.1.8	Research Instrument 2: Portfolio of competence	84
5.1.9	The case for looking at reflection	84
5.1.10	Research Instrument 3: Reflectivity questionnaire	85
5.1.11	The case for looking at consulting	85
5.1.12	Research Instrument 4: Patient Enablement Questionnaire	86
5.2	Conceptual Considerations around Analysis: MLM in Education Research	86
5.2.1	Options for Statistical Analysis	88
5.2.2	MLM working in parallel situations to my own research	88
5.2.3	Multilevel Models for school and teacher effectiveness	89
5.2.4	Multilevel Theories	90
5.2.5	Classes of Multilevel Models	90
5.2.6	Hierarchical nature of educational data in MLM	91
5.3	Choices to be made in my research	91
5.3.1	Is MLM the best approach to adopt?	93
5.4	Designing the multilevel model	93
5.4.1	Considerations regarding levels	93
5.4.2	Designing the multilevel model: Response Variable	94
5.4.3	Designing the multilevel model: Explanatory Variables	94
5.4.4	Designing the multilevel model: Complex variation at level one	95
5.5	Measurement in Phase Three	95
5.5.1	Recruitment problems	95
5.5.2	Insight 360 Perceptions	95
5.5.3	Responses to questionnaires	96
5.5.4	Examining the data	97
5.5.5	How do perceptions vary?	98
5.5.6	Factor Analysis	101
5.5.7	Looking at the data with MLM	102
5.5.7.1	Doctors' self-perception in 'Patient Care'	104
5.5.7.2	Importance and Performance Ratings	109
5.5.7.3	MLM analysis of self-perception in the other categories	110
5.5.7.4	Patient perceptions analysed by MLM	111
5.5.7.5	Colleague Perceptions	115
5.6	Reflection: Phase Three	116
5.6.1	Reflection on Phase 3 recruitment problems: the constraints on GP time	116
5.6.2	Reflections on Phase Three measurements in Insight 360	117
5.6.2.1	Questionnaires	118
5.6.3	Reflection: How well do Phase Three findings accord with theory surrounding perceptions?	118
5.6.3.1	What are the determinants of self-perception, other than performance in the tested domain?	118
5.6.3.2	Self-perceptions and gender	118
5.6.3.3	Why is our data so compact?	119
5.6.3.4	Do perceptions still reflect meaningful dimensions of quality?	120
5.6.3.5	What are the other biases in Person Perception?	121
5.6.3.6	Consistency of Patient Perceptions	121
5.6.3.7	Speculating about individual perceptions	122

Evaluating Process and Outcome in the Education of General Practitioners

Table 5.1	The Research Programme in context of General Practice Scene	81
Table 5.2	Examples of evidence of assessed competence, which could be used for a portfolio either for revalidation purposes or for use as quality indicators in research	83
Table 5.3	Information requested for Portfolio of Competence	84
Table 5.4	Techniques that can be used to look at any level of an educational process	87
Table 5.5	Advantages of multilevel models	88
Table 5.6	A scheme for levels in MLM	92
Table 5.7	Planned explanatory variables in my research	94
Table 5.8	Insight 360 categories	96
Table 5.9	Non-response items	96
Table 5.10	Data reliability	97
Table 5.11	Perception scores for a possible outlier: doctor 15	99
Table 5.12	Perception scores for a possible outlier: doctor 22	100
Table 5.13	Gender influence in the four categories	110
Table 5.14	'Importance' influencing assessments of 'Performance' in the four categories	113
Table 5.15	Biases and Problems in Person Perception (after George, 2002)	121
Table 5.16	Legitimate uses for perception-based multi-source performance data	123
Chapter 6	Reflections On The Research Project	
6.1	The original intentions reviewed	124
6.1.1	What has been achieved?	124
6.2	Categories and Dimensions of Educational Behaviours	124
6.2.1	Scope for refinement of categories of Educational Behaviours	125
6.3	The assessment of educational behaviours in GP training practices	126
6.3.1	Overcoming problems of documenting behaviours on visits	126
6.3.2	Changing educators' behaviours	127
6.3.3	The Clinician Educator as Role Model	127
6.3.4	Committees sitting in judgement	128
6.4	Perceptions and multi-source feedback data	128
6.4.1	Formative Use of Perceptions and feedback	130
6.4.2	Potential for 360 degree perspectives feeding into the revalidation process	130
6.5	Method of analysis	131
6.5.1	Will it ever be possible to relate educational process to outcome?	131
6.6	The way forward – further research	132

References	133
Appendices	
1 Glossary: Acronyms, Abbreviations and Definitions	141
2 Conceptual Model: A Trainers' Performance	143
3 Categories and Dimensions of Educational Behaviours	155
4 Trainers information about assessment visit	156
5 Team-leaders' plans for Phase IIc	157
6 Team-leaders' study day revising framework	165
7 Attributes of an assessment process	174
8 Domains of assessment for quality in General Practice	175
9 Insight 360 patient questionnaire	176
10 Detail of Insight 360 questions not attempted	181
11 Report from Vanessa Simonite	184
12 Further results of multilevel model	191
13 Standards for quality measure	203
14 Semi-structured interview template (revised after pilot)	204
15 Reflectivity questionnaire	205
16 Self-assessment of educational Behaviours	207

CHAPTER 1
INTRODUCTION
Thesis Structure And Theoretical Issues

Every one of the links between what we teach, how this is translated into performance at work, and how this performance influences the quality of health care, let alone the quality of health, remain leaps of faith.

(Marinker, 1992)

1.1 Structure of the thesis (“Concept”, “Measurement”, and “Reflection”)

This thesis describes my work to develop a methodology relating an important part of the process by which novice practitioners learn their craft to the outcome in terms of their subsequent performance in practice.

To engage with this work it is necessary to understand something of the educational issues, which I shall cover in this chapter, and something of the situational context of General Medical Practice, including the educational structure, which forms the subject of the next chapter.

At the outset, it is necessary to understand that intending General Practitioners (GPs) spend a year as a ‘Registrar’ in a training practice, where their vocational learning is a priority. They undertake medical duties, including a large number of consultations with patients, under the supervision of a GP ‘Trainer’, who is a doctor trained and accredited to teach one-to-one throughout the training year. How this teaching and learning happens, and how effective it is in the long term, is the subject of my research.

My work was designed and executed in three phases. Phase One researched the different processes by which Trainers help GP Registrars learn. Phase Two looked at how such educational processes can be assessed in training practices. Phase Three looked at possible outcome measures in practice.

Each phase is discussed in a separate chapter, and a uniform format is adopted for each chapter, whereby I look first at the conceptual issues around study design, I then

Evaluating Process and Outcome in the Education of General Practitioners

describe what measurements I have undertaken, before finally reflecting on the process.

1.1.1 Thesis Overview

This format, which can be abbreviated as Concept, Measurement, and Reflection, is mirrored in the macro-design of the thesis:

- Concept Chapters (Chapters One and Two)
- Measurement Chapters (chapters Three, Four, and Five)
- Reflection Chapter (Chapter Six)

1.1.2 Concept Chapters

Two chapters are concerned with setting the scene and with the conceptual underpinning. The first explores the nature of the evidence and the judgements to be made on the basis of such evidence. Key concepts such as performance and competence; appraisal and assessment are explored and defined. The second chapter contextualises these concepts in the world of medical education for GPs.

1.1.3 Measurement Chapters

The next three chapters detail the measurement phases of the study, and the reader needs to be aware that these chapters are quite different one from another as they engage with totally different methodologies across the qualitative and quantitative paradigms.

Chapter 3 describes interviews based on interpersonal recall and a grounded theoretical approach to evidence about the significance of interactions with a GP Trainer.

Chapter 4 is more about judgements based on evidence. These are the judgements that are made about the quality of education in training practices.

Evaluating Process and Outcome in the Education of General Practitioners

Chapter 5 also concerns evidence and judgements; this time it is the evidence around quality of practitioner, and the synthesis of multi-source perceptions. Here the reader is asked to engage with statistical multi-level modelling (MLM) techniques.

Within these 'Measurement Chapters' the micro-structure is again: concept, measurement and reflection. The available methodologies are first discussed; the chosen methodology in action is detailed, and the performance reflected upon.

1.1.4 Reflection chapter

The final chapter is reflecting on the methodologies, which I have evaluated. What has my work achieved and what should be the direction of onward travel? What are the pointers towards the further work, which needs to be done.

1.2 What is so elusive about process and outcome?

Most medical treatments require evidence of an outcome of patient condition improvement to warrant national investment (Southgate, 1994). Likewise, there is a drive to seek evidence that educational initiatives affect learner outcomes (Ashley, 2000), yet it is unusual to find medical education programmes that are funded on the basis of solid evidence of outcomes improving patient care.

Alarmed by the lack of evidence to support the judgements made at times of re-accrediting educators in General Practice (Peile and Johnson, 2002), I resolved to try to develop a methodology for relating process measures to outcome in the nationwide programme for training new GPs. In particular, I wanted to look at the interaction between Trainer and Registrar, during the yearlong period of one-to-one attachment for training. Vocational Training is an expensive process, and we need to look at best value for money.

Some knowledge of the context of this area of medical education is important to understanding my work, and this will be developed in the next chapter, but the reader is also referred to the glossary and definitions in Appendix 1.

Evaluating Process and Outcome in the Education of General Practitioners

Throughout this chapter, the focus of attention will oscillate between the educational practice of the doctor as Trainer (the process I research) and the clinical practice of the doctor as caregiver (the outcome in which I am interested). It is my contention that the quality framework for defining practice, for recording perceptions, and for assessing and improving practice is essentially similar across both educational and clinical domains of practice.

There is potential benefit from parallel process: in an already over-complex professional environment, doctors are likely to operate better in a familiar framework. In seeking to clarify some of the concepts around perceptions, I constructed a series of very simple conceptual models (see Appendix 2).

1.3 Performance, Perceptions, and Assessments in General Practice Education

Central to this work is a consideration of performance and competence of doctors, both as clinicians and as educators, so it is worth defining terms and looking at a conceptual interrelationship.

1.3.1 The Good Doctor

The General Medical Council (GMC), currently introducing regular mandatory performance review, charges all doctors to keep professional knowledge and skills up-to-date and to “recognise the limits of your professional competence”, as “All patients are entitled to good standards of practice and care from their doctors” for which professional competence is essential (GMC, 2001). Establishing what is “good” is problematic. Pringle and colleagues (2002), in an article entitled ‘Measuring “goodness” in individuals and healthcare systems’, admit the inadequacy of the measures used, and the need for value judgments in arriving at a conclusion.

1.3.2 Performance as distinct from Competence

Some educationalists, taking a stance based on assessments, view performance as how a pupil performs in an examination, and competence as what the pupil is capable of achieving. I, on the other hand, adopt Grol *et al*’s (1989) definition of performance

Evaluating Process and Outcome in the Education of General Practitioners

as ‘what a doctor actually does in daily practice’, as opposed to competence, which represents what he or she is capable of doing in ideal circumstances.

Performance is what is of interest in maintaining professional standards (Irvine, 1997; GMC), but competence is what is most frequently assessed, both at undergraduate and postgraduate level (Van der Vleuten, 1996; Pitts, Coles *et al.*, 1998). Competence can be inferred from a doctor’s performance.

Whether one views performance as *performance in tests* or as *performance in practice*, competence is likely to exceed performance: the doctor at work and the student in the examination are both unlikely to perform up to the limits of their competence.

Again the two views are similar in that possession of the requisite competence can be inferred from good performance, but poor performance does not necessarily infer a lack of competence (one may be capable of performing better, but not do so).

However the two views differ in their longitudinal perspective. *Performance in tests* requires but a brief focussed achievement, whereas *performance in practice* reflects sustained effort, and any measurement of this invokes concepts of sampling over time and averaging, if we are to look beyond measures of peak performance.

1.3.3 Performance over time

There are a number of considerations about performance, as I define it. Not least is the level of constancy in clinical practice. The evidence is conflicting (Cunnington *et al.*, 1997; Pringle *et al.*, 2002), but supports the assertion that GPs perform to a relatively constant level, allowing for peaks and troughs (Burrows *et al.*, 2001). As Pringle *et al.* (2002) point out, not all differences are meaningful, and the key task is to detect unacceptable or dangerous variation in performance indicators.

Likewise the evidence on medical teachers, albeit from undergraduate education, supports a relatively constant level of performance (Dolmans *et al.*, 1996).

What is less clear is the level of context specificity in teaching performance (Combleth, 1991).

The best guide to how a doctor will perform clinically is how they have performed previously (Wood and O'Donnell, 2000), and to ascertain this, it is worth seeking examples of recent past practice (Hayden and Adams, 2000).

1.3.4 Competence

Without the necessary competence it is impossible to perform at the required level, but even the competent may fall short of the required level of performance.

‘Competence is concerned with what people can do rather than what they know’ (UDACE, 1989). It is contextual; an outcome rather than a process measure of education; and does not necessarily endure. Competence across a broad area of practice invokes achieving a number of competencies – the atomic components of competence, which can be more easily defined and measured than the integrated competence. The reductionist nature of competency-based training has been well exposed (Hyland, 1994), and I would argue that meaningful assessments of professional education of generalists must grapple with the less easily measurable, but more relevant holisms of competence that predicate performance as a doctor.

Important domains of professional competence, such as the integration of knowledge and skills, the context of care, information management, team-working, or patient-doctor relationships often escape assessment scrutiny (Epstein and Hundert, 2002).

Medical competence is multidimensional. In the definition generated by Epstein and Hundert (2002), who conducted a systematic review, the following elements were included:

... communication, knowledge, technical skills, clinical reasoning, emotions, values, and reflection in daily practice for the benefit of the individual and the community being served.

Dutch GPs, who had completed Vocational Training in one university 10-15 years previously, varied little in their knowledge of somatic aspects of the consultation, but varied considerably in their degree of patient orientation and their risk-appraisal. Furthermore, it seemed that the least competent were also the least prepared to consult their colleagues about diagnostic and therapeutic problems (Kuyvenhoven *et al.*, 1990) and were thus likely to continue to perform less well.

McGuire (1983), in a thorough review of the research on evaluation of professional competence, discusses the merits of simulation versus observation of actual performance, coming down on the side of observing actual performance. More

recently, McKinley *et al.* (2001) advocated that direct observation of a GP consulting should always form a part of assessment. They see direct assessment of competence and indirect performance review as “complementary”. At first this statement may appear at odds with my definitions of competence and performance, but in fact it is concordant. A practitioner consulting under observation will take care to demonstrate that his/her skills match the published competency framework. The observed consultation shows what the doctor can do. We need also to know what the doctor does – hence the emphasis on ‘indirect performance review’ throughout this thesis.

1.3.5 Perceptions in assessment

Few assessments observe trainees in real-life situations; incorporate the perspectives of peers and patients; or use measures that predict clinical outcomes (Epstein and Hundert, 2002). Mindful of the paucity of such data, I determined to include multi-source data in this work. Multi-source data is essentially an amalgamation of people’s perceptions, so it is important to examine the legitimacy of amalgamating perceptions. The case for such data, in supporting critical judgements on performance, was laid out in an international review on poorly performing doctors (Southgate, Cox *et al.*, 2001b). I engage in a discussion of what determines perceptions in Chapter 5, when I am reflecting on the Phase Three results. In the meantime, I constructed a working definition of a perception on which a professional assessment is made: an opinion, based on experience, and interpreted subjectively, consisting of more-or-less subconscious judgements about performance.

In the next chapter, I explore some issues of perceptions specific to the medical context. I look at the perceptions of medical learners about the training process, at some of the issues specific to patient perceptions and to the perceptions of medical colleagues.

1.3.6 Appraisal and Assessment defined

I adopt the definition of appraisal as a formative review of performance and competence, for the benefit of the appraisee, resulting in the setting of educational objectives (DoH, 2002). By contrast, assessment, which may also be formative for

Evaluating Process and Outcome in the Education of General Practitioners

the doctor assessed, has summative dimensions, and is owned by the assessor or the assessing organization. Appraisal benefits the organisation, ensuring that doctors are working to the organisation's objectives, as well as ensuring that public expectations of standards are met. Both appraisals and assessments involve judgments, and unless appraisers and assessors are able to directly observe representative practice, both processes depend on input involving perceptions. In the case of appraisal, there is more emphasis on self-perception, but assessment also benefits from a judicious balancing of self-perception with the perceptions of other stakeholders.

I have more to say about appraisal and assessment in medicine in the next chapter, which includes assessments of medical teachers, as well as assessments for the purposes of revalidation of doctors.

1.4 Actions, Evidence about them, and Judgements upon them.

For reasons, which are discussed in detail in Chapter 5, much of the **evidence** about quality of performance of doctors as educators and clinicians is based on **perceptions**. **Judgements** are made on the basis of such evidence: we need therefore to look at this conceptually. I found it helpful to construct a conceptual model to guide my thinking, by breaking down the process into different stages. My model is reproduced at Appendix 2. Reflecting on this, I detail the following considerations, at the levels of action, evidence, and judgement.

1.4.1 The actions, which are the subject of report

All evidential perceptions and judgements are focussed on the subject's clinical or educational activity, which constitute the actions under report. There are some considerations that affect the ways in which the action may be perceived.

Transparency - patients and (to a lesser extent) learners may only be able to observe parts of the doctor's performance. In the selection of a therapeutic or educational approach, much skilful weighing up of alternative options may have proceeded inexplicitly. (As in the phenomenon of the swan appearing to glide effortlessly upstream, when the legs are paddling hard under the surface.)

Evaluating Process and Outcome in the Education of General Practitioners

Representativeness - is the observed action representative of the doctor's performance or atypical? Is the doctor having a bad day or putting on a special performance?

Cross-referencing - could observers' perceptions of the GP Trainer's performance as a doctor have coloured observations regarding performance as an educator? There is much in common: communication skills and an ability to reflect purposively are examples of qualities common to both domains of performance, and I propose parallel quality frameworks, yet an able clinician can be a poor educator and vice-versa.

There are two secondary, possibly opposing, considerations influencing perceptions of the action. They are the outcome and, at the other pole, the original motivation.

Is the action being evaluated by outcome rather than execution? Well-chosen actions may be ineffective, whilst injudicious actions may have successful outcomes.

Perceptions may be based on outcome (*I got better*) rather than process; (*the antibiotic treatment for a viral illness cost me money and exposed me to unnecessary risk.*)

By contrast, motivation may be a particular confounder for self-perception, if the actor looks at his/her intent rather than performance. Others may be misled about the doctor's concern for the patient/learner's interests. (Dr. Harold Shipman, whose malevolence was concealed, was well liked by many of his patients and colleagues.)

1.4.2 Evidence of observers

As will become apparent in Chapter 5, I am interested in the perceptions of three different classes of observer. These are respectively the subject doctor, reporting his or her own self-perceptions on educational or clinical activity; the learner or the patient who interacts with the doctor in those activities; and the colleagues, who working closely with the doctor, have insight into the quality of performance.

1.4.2.1 Perception formation

There are a number of primary considerations about how observers form their perceptions in my opinion. Contextual Influences are pre-eminent. Sen (2002) talks

Evaluating Process and Outcome in the Education of General Practitioners

about the patient's self-perception of illness in a social context being affected by 'levels of education, availability of health facilities, and public information on illness and remedy'. Related to this is way in which different individuals filter complexity – a multidimensional attribute may be perceived through a single filter which picks out selected attributes. For example the patient may care more about the doctor's kindness and listening skills than about technical competence.

Cultural considerations for patient care medicine extend beyond dimensions of faith, ethnicity, age, and social class, to those of beliefs around conventional and alternative medicine. Likewise the diversity of learning styles introduces numerous cultural dimensions into education.

Another consideration is proximity. A particular concern regarding colleague observations is the ability of the colleague to get close enough to observe the subject's actions. As consultations and tutorials take place behind closed doors, the perceptions of colleagues are often heavily dependent on second-hand observations: for example, hearing patients' opinions, seeing referral letters and investigations.

Are multi-source perceptions independent of each other or do patients' perceptions depend to a significant extent on subliminal messages from receptionists and other colleagues of the doctor? Conversely, do colleagues base perceptions on their own observations, or are they reflecting the doctor's reported popularity with patients?

Are all perspectives equally relevant, and should we weight them all equally? Is self-perception relevant to all aspects of the doctor's activity? Are there activities for which self-perception is the most relevant perception? Which are the activities, which can be meaningfully reported on by patients? Are all colleagues able to report on all activities? These are questions, which can only be answered by collection, analysis, and discussion of multi-source data.

1.4.2.2 Perception recall

Memory fallibility is a real concern when reporters are asked for their perceptions of temporally distant interactions with a subject. A particular form of recall bias happens

Evaluating Process and Outcome in the Education of General Practitioners

as the passage of time applies a filter of relevance – only significant encounters may stand out at a distance.

Although our personal constructs are reasonably stable in the short-term, there is evidence of considerable long-term shifts (Kelly, 1955). Thus, at the time a learner or a patient experienced an interaction with the subject, it might have been perceived positively, but subsequent experience may have coloured opinion, so that in later years the experience is reported negatively.

In chapter 4, I will discuss the influence of committee discussions on individual opinions. Perceptions of the care offered by a GP are influenced by conversations around the dinner table or in the supermarket, in much the same way as processes such as sharpening or levelling of perceptions (Cicourel, 1976) happen on day release courses for GP Registrars.

If patients' perceptions are requested in the waiting room, the 'cues' of the surgery, permeated by the receptionist's voice and internal feelings of discomfort, may help or hinder meaningful reporting of past encounters.

1.4.2.3 Perception transmission

For judgements to be made on the evidence, perceptions have to pass from observers to assessors, and this process usually involves intermediary steps. Perceptions can be collected by means of questionnaires, by interviews, or even by spontaneous reporting. Whichever process is involved, selective 'filters' introduce bias, often systematically, into the process. Questionnaires ask structured questions, and like interviews may 'lead' the witness, whereas spontaneous reporting favours extreme examples.

In the process of interpreting perceptions, the language constraints of perception recording and analysis may lead to systematic or random biases at the level of analysis – "*What you think I said is not what I meant to say*".

1.4.2.4 Aggregation of perceptions

There is a tension in that the higher the number of observations, the greater the confidence we can have that the prevailing perception is representative of the action under report – that we are not looking at evidence of freak behaviour which is unrepresentative of actual performance.

The evidence from 360 degree perceptions is that ‘saturation’ is achieved by 15 perceptions from any one group (Griffin, 2000). Collecting a large number of perceptions across a large number of activities for a large number of doctors makes considerable demand on data-handling, and a process of aggregation is inevitable.

1.4.3 Judgements

How accurately can assessments of performance be made on the basis of perceptions of others about experiences, which happened some time ago in another place? Those responsible for revalidation of doctors and for accreditation of Trainers, are basing those judgements on evidence of perceptions based on remote experiences. Is this process reliable? These considerations are crucial to Chapter 4 and Chapter 5.

Appendix 2 shows how a simple model may be built up to address these considerations which underline the work of this research project, describing work using perceptions to address issues of quality in clinical practice and medical education. I shall return to these issues in Chapter 6.

For all the difficulties surrounding the making of judgements, it is my contention that a large part of the reason for the paucity of evidence linking process to outcome in GP

Registrar education is attributable to the difficulty of agreeing meaningful categorisation of process and meaningful measures of outcome. This is the rationale for my attempts to develop appropriate methodology

CHAPTER 2
THE RESEARCH CONTEXT
GP Training

2.1 GP Registrar training

A minimum of one year spent as a trainee in an approved training practice has been a legal requirement for general practice since 1979 (NHS Vocational Training Act 1976). To become a GP in the United Kingdom, qualified doctors, (registered with the GMC) must first undergo at least three years of postgraduate Vocational Training. Parliamentary Regulations specify the length and content of Vocational Training, which is currently centrally controlled by the Joint Committee on Postgraduate Training for General Practice (JCPTGP). The training is organised on a deanery basis, in Departments for Postgraduate Medical and Dental Education (PGMDE). Responsibility for the provision of training within each Deanery rests with the Director of Postgraduate General Practice Education, supported by a deanery team of Associate Directors. In each deanery, there will be a number of different Vocational Training Schemes, (VTS), each centred on a local Regional or District General Hospital, and directed by one or more local Course Organisers.

The training programme usually consists of 24 months in approved hospital posts and 12 months in general practice as a GP Registrar. Doctors who wish to train for general practice usually apply for a place on a three-year VTS. They can also apply for recognition of a self-selected programme made up of posts in the specialities and locations of their choice, provided these posts are approved for general practice training.

The time spent in General Practice, supervised by an approved GP Trainer, is the subject of my work. Some schemes (including those in HM Forces) permit 18 months in practice, and the time can be split, but it is usual that the final phase of the training programme is spent as a GP Registrar for a period of at least six months. This allows the experience of hospital based training to be the subject of subsequent reflection in a General Practice context. Some GP Registrars experience training in two separate practices, under different Trainers, but it is more usual to train under a single Trainer. The practice in the Aylesbury VTS, where the Phase One work took place, is for

Evaluating Process and Outcome in the Education of General Practitioners

Registrars to spend an initial month with their Trainer, before returning to hospital for two years rotating through relevant junior posts in specialities such as Paediatrics, Obstetrics, Psychiatry, or Care of the Elderly. These doctors then return to their training practice for the final 11 months of their formal training.

All entrants to General Practice have to undergo Summative Assessment, the four components of which are designed to test that they have acquired the relevant knowledge, skills and attitudes for independent practice. This compulsory assessment is designed for the General Practice setting.

Recruitment to both formal schemes and stand-alone GP Registrar posts is co-ordinated centrally by each deanery. GP Trainers are no longer permitted to recruit directly to posts in their practice, but the Trainer would have selected those in my Phase One study, as was then the practice.

2.2 General Practice Trainers: Accreditation And Development

Training practices enjoy financial rewards in the form of a Trainer's grant, in addition to the service commitment of the GP Registrar, which is cost-free to the practice, so it is proper that the training system is well regulated and accountable.

The quality of teaching offered by a GP Trainer is likely to impact on the GP Registrar, and the presumption is that this will affect the quality of the Registrar as a doctor proceeding into practice as a GP. It is also presumed that the model of doctoring demonstrated by the Trainer is important to the GP Registrar's formation, and this (somewhat undertested) premise underlies the insistence that training practices should demonstrate high standards of clinical practice.

2.2.1 Accrediting the Trainers nationally

Postgraduate medical education was, until relatively recently, poorly regulated and of variable quality. Indeed, the same might be said of undergraduate teaching when, ten years ago, only 19% of teachers at a British medical school had attended a course in medical education in the previous five years (Finucane, Allery *et al.*, 1992).

In 1952, the year the Royal College of General Practitioners (RCGP) was founded, some Inverness doctors started the first scheme where General Practice traineeships

Evaluating Process and Outcome in the Education of General Practitioners

were linked to relevant hospital posts. By 1959 an experimental training scheme for General Practice had started in Wessex, and by 1972, when the first Regional Advisers in General Practice were appointed, there were about 300 trainees and Trainers in the UK. A period of rapid expansion followed over the next decade and the number of trainees and Trainers rose to 2,750 and 2,000 respectively. Today there are approximately 3,500 GP Trainers in the UK (JCGPT).

The JCPTGP, which started in 1976, acquired responsibility under the National Health Service (Vocational Training) Regulations 1979 to exercise general oversight of the training standards throughout UK and a statutory function to inspect training schemes. It could not, however, remove approval from individual Trainers or hospital training posts until the NHS (Vocational Training for General Medical Practice) Regulations 1997, gave JCPTGP the responsibility, by law, for the approval of all posts used for GP training both in hospital and in general practice (JCGPT, 2001).

The 1994 Vocational Training for General Medical Practice (European Requirements) Regulations appointed JCPTGP as the Competent Authority for General Practice training under European Law ensuring that the training process in all deaneries meets European standards as laid down by the Directive (Council Directive 93/16/EEC). JCPTGP now organise a triennial peer review visit to examine a deanery's own performance in monitoring the delivery of its education to its Trainers and trainees. The visit also serves as a method to formally accredit the deanery's GP vocational training for a period of three years (JCGPT).

Improving the evaluation of faculty teaching undoubtedly remains one of medical education's greatest challenges (Jones and Froom, 1994).

2.2.2 Accrediting the Trainers locally: peer review visits

Even before JCPTGP laid down the standards, deaneries developed their own selection and approval criteria for training practices, and the Oxford Deanery piloted and refined the peer review visits to General Practice Trainers, which are the subject of my work. Over the past 20 years, such visits have been made to all Trainers, at a minimum of 4-yearly intervals, by teams of two GP training peers and a Team-leader

Evaluating Process and Outcome in the Education of General Practitioners

(Schofield and Hasler, 1984; Schofield and Hasler, 1984; Schofield and Hasler, 1984). Team-leaders are drawn from the pool of GP educationalists, and may be: another experienced Trainer, a course organiser or an Associate Adviser, a Deputy Director or the Director of General Practice training. Since 1995, teams have also included a Practice Manager from a training practice, who has contributed to the assessment of practice organisation in the visited training practice (Johnson, Hasler *et al.*, 1997).

2.2.3 Accrediting the Trainers locally: Team-leaders' reports

The visiting team, who are sent a portfolio of evidence about the training in advance of the visit, spend a whole day visiting the practice. They then contribute to a report, written by the Team-leader, which is intended to reflect a synthesis of the observations of the visiting team.

The criteria for training practices are set by the Deanery, in line with national guidelines. In advance of the team's visit, all established and prospective Trainers have been asked to supply evidence that these criteria are met. To a large extent, the function of the visiting team is one of verification that both Trainer and practice meet the required standards. However, the visit also affords opportunities for formative development, and the team uses their combined experience to make recommendations about how the training might improve. Visiting teams are further encouraged to commend exemplars of good training practice, and to express their reservations about training, which although meeting the criteria, may seem to have some undesirable features. Team-leaders' reports are therefore qualitative documents, which include much that is objective alongside some subjective judgements.

2.2.4 Accrediting the Trainers locally: Trainer Selection Committee

The Team-leaders' reports are presented to, and considered by, the Trainer Selection Committee, which meets quarterly to consider all doctors who are applying for accreditation or re-accreditation as GP Trainers. The Committee, chaired by the Director of Postgraduate GP Education, then decides whether to approve the doctor and the practice for training GP Registrars.

Evaluating Process and Outcome in the Education of General Practitioners

It is made clear that visiting teams do not decide the outcome of a Trainer's application – the Trainer Selection Committee decides this. The main source of evidence that the committee considers is the Team-leader's report, and members of the committee are discouraged from bringing prior knowledge to bear on their decision-making process. The Committee comprises the Director of Postgraduate GP Education, his Deputy, an elected member of the local RCGP Faculty Board, an elected representative of GP Registrars, a representative of Deanery Trainers, the Team-leader for the visit, the local Course Organiser, and the Local Medical Committee (LMC) representative. The judgements made in this Committee are examined in depth in Chapter 4.

2.2.5 Educational needs of GP Registrars - what is the evidence?

Vocational Training is an expensive process, which merits critical appraisal. Since the work of (Freeman and Byrne, 1976) there have been other attempts to look critically at Vocational Training and at Continuing Medical Education (Marinker, 1992; Smith *et al.*, 1998). Despite some very erudite analysis (RCGP, 1985; Calman, 1994; Southgate, 1994), a lot remains elusive about the process of learning and what facilitates the development of excellence in GPs.

2.2.6 Evidence about the impact of GP training on subsequent performance in Practice

We know that knowledge increases during vocational training to a peak, which is maintained for the first ten years of postgraduate practice, and which then declines thereafter (Van Leeuwen *et al.*, 1995).

Although some qualitative work shows that education plays only a relatively small part in influencing doctors' behaviour (Smith *et al.*, 1998), there is evidence that vocationally trained GPs are

better GPs in terms of performing the tasks of a GP to a level of providing quality care. (Hindmarsh *et al.*, 1998)

Hindmarsh's group carried out a review of the international literature reporting outcomes of general practice vocational training programmes. Twenty-five studies were reviewed, including five that draw on learners' or teachers' open-ended accounts of the impact of the vocational training programme. None of these specifically

Evaluating Process and Outcome in the Education of General Practitioners

focuses on the learners' evaluation of the Trainer and training practice. Shapiro and Talbot (1991) report that teachers of GP Registrars particularly noted the impact of the programme in relation to increased confidence in Registrars dealing with uncertainties. Training also increased knowledge of the processes of general practice - decision making, conceptualisation and interaction with patients.

From the evidence available, Hindmarsh *et al* (1998) concluded that:

The studies reviewed suggest that vocationally trained GPs are 'better' GPs in terms of their quality of patient care, confidence and self perceptions as a GP, knowledge base, specific practice skills, attitudes and personality traits, adherence to practice guidelines, medical intervention patterns and examination pass rates. No studies indicated that vocational training had negative effects on practice or no effect at all.

After Hindmarsh *et al* (1998) published their multifaceted summary of evidence, other studies have endorsed the conclusion that Vocational Training is effective, such as the study of 212 doctors from the West Midlands which has shown that 12 months training as a GP Registrar has an important impact on the development of perceived skills in palliative care (Charlton, Field *et al.*, 2000). No studies have yet shed light on the processes that contribute to improved performance by learners.

A lot of attention has been placed on the day release component of training and what the scheme as a whole offers, for example Grol *et al* (1989) have compared the impact of a systematic training in consultation skills against a problem-based learning approach. By contrast, there seems to have been little work published, which looks at the perceived value of different Trainers and different training practices. Thirty years ago in a report entitled 'The Future General Practitioner', the RCGP observed, "By questioning our thinking and our practice the trainee makes us look at ourselves" (RCGP, 1972). It seems that little use has since been made of the learner as a mirror.

2.3.1 Perceptions In Medicine and Medical Education: Perceptions of ex-trainees

Sean Hilton (1981), in his personal reflection on a trainee year comments that there are doctors who:

...might argue...that all facets of general practice...will anyway be absorbed by osmosis within the first two years in practice, and that the only way to learn real general practice is to get on and do it.

There have been many attempts since then to evaluate the usefulness of vocational training. Kelly and Murray (1991a, 1991b) looked at twenty years of vocational training in the West of Scotland, primarily from a structural viewpoint, and comment that their study

...emphasises that the opinions of ex-trainees are a worthwhile and under-utilised source of information which is of great importance to all bodies involved in vocational training.

Trainees who have finished their training are ideally placed to give useful feedback on their training and Trainers, but their comments must be considered in an appropriate context (Brahams, 1982).

Anyon (1987) undertook a qualitative evaluation of the first 10 years of the New Zealand family medicine training programme. As part of this study, ten randomly selected past Registrars were asked about their experience of the impact of the programme on their practice. All said that they'd learnt a lot about General Practice in the widest sense, although they considered some specific subjects had been poorly addressed.

Published surveys of trainees' views on their training year have usually been carried out at the conclusion of training, and none has addressed what has added value to lifelong learning in General Practice (Whitfield, 1966; Freer and Reid, 1978; Martys, 1979; Thornham, 1980; Hilton, 1981; Ronalds, Douglas *et al.*, 1981; Anyon, 1987; Short, 1987; Crawley and Levin, 1990; Duncan, 1994). Only when trainees have been working as GPs for a considerable length of time, are they in a position to provide reflective feedback on which elements of their training have proved the most worthwhile in practice.

2.3.2 Perceptions In Medicine and Medical Education: Patient perceptions

Richard Baker (1997) points out that essential to measuring and interpreting patient satisfaction is a theoretical basis to explain the meaning of satisfaction, and hence how it should be measured and how the findings are interpreted. He talked about developing a 'pragmatic' model linking together

...empirical evidence about patient satisfaction without recourse to more general social or psychological theory of behaviour, other than to define satisfaction as an attitude.

Early work showed a need to adjust the questionnaires to reflect the importance of personal care to patients. This is an example of 'satisfaction' as a perception in terms of the above definition.

As McKinley *et al* (2001) point out, high levels of patient satisfaction, cannot, on their own, be relied on to indicate competence, nor low levels a lack of competence. They cite the example of a patient who may be dissatisfied with the professionally correct refusal to agree to an inappropriate request for hypnotics or antibiotics.

2.3.3 Perceptions In Medicine and Medical Education: Perceptions of colleagues

McKinley *et al* (2001) point out that the views of colleagues may not always truly reflect performance, as negative feelings among peers may reflect problems in the professional relationships of a doctor who still manages to provide good care.

I would, however, maintain that as effective team-working is integral to General Practice, an essential dimension of performance is the nurturing of good professional relationships – indeed the GMC reinforces this in 'Good Medical Practice' (GMC, 2001). The perceptions of colleagues are indispensable to the assessment of a doctor's professional relationships. Given that colleagues may also be capable of being professionally dispassionate, it is likely that their opportunities for close observation permit valuable perceptions on a doctor's clinical practice and teaching.

Who are a GP's colleagues? Most doctors work in groups or partnerships, and their relationship with their fellow doctors on-site may be one of legal partnership (with or

Evaluating Process and Outcome in the Education of General Practitioners

without hierarchy of seniority) or on either side of an employment relationship, as Assistants, Retainees, and indeed Registrars are usually employed by their colleagues.

Examples of doctors' windows into their colleague's **clinical** practice include:

- Patient feedback: *"I don't feel comfortable talking to Dr Smith about this", or "Dr Jones was so caring when she looked after me when you were away"*
- Records: the quality of clinical notes, the ease with which it is possible to follow another's care plan, the appropriateness of clinical investigations and referrals.
- Discussion: opportunities to discuss a topic of clinical interest arise almost every coffee break, and colleagues may perceive the strengths and weaknesses of another's knowledge base. There is even evidence to suggest that unwillingness to consult colleagues may be a marker of low competence (Kuyvenhoven, Pieters *et al.*, 1990).

Examples of doctors' windows into their colleague's **educational** practice include:

- Learner feedback: *"Dr Smith gave me a brilliant way of looking at depression"* or *"Dr Jones doesn't seem to like discussing cases with me"*
- Records: the Trainer's educational planning tools are often visible to colleagues, as is the Registrar's log.
- Discussion: others besides the Registrar may experience teaching from a doctor. A colleague will have insight into the facility with which the doctor can tease out another's difficulty and explain difficult concepts.

However other colleagues may also have perceptions about a doctor's clinical and educational practice. Changes in the pattern of skills-mix often mean that a GP works more closely with nursing colleagues than other doctors, as doctor and nurse often share responsibility for a clinic or for particular aspects of care. Joint-teaching is increasingly a feature of the modern teaching practice, and colleagues from other disciplines have valuable insights not only into whether their professional skills are appropriately used in cross-referral, but also whether they are optimally used educationally.

Evaluating Process and Outcome in the Education of General Practitioners

Likewise, receptionists and office staff, as the practice interface with the public have valuable perceptions to offer. They usually know more about a doctor's popularity, and the reasons underlying this, than other practice team-members. In the (often unintentional) hierarchy of Primary Care Teams, the Registrar may find the most comfortable sounding-board in the back-office!

2.3.4 Perceptions In Medicine and Medical Education: Self-perception

The importance of self-perception is threefold. Firstly, no one else has such broad experience of the doctor's practice as the doctor in question. Secondly, to encourage active self-perception is to encourage a process of reflection-in-practice and reflection-on-practice, to the betterment of professional practice (Schön, 1983).

Thirdly, self-perception if not overt, remains subconscious: it may be the misalignment of this subconscious perception with the perceptions of others that leads the doctor to reject their input. The more healthy process is to bring self-perceptions to the surface and use them as a template against which to reflect on the perceptions of others. This is the theory behind 360-degree appraisal (King, 2002).

There is evidence to suggest that GP's self-assessment of knowledge may be seriously 'flawed' as determined by objective testing, and triangulation is therefore highly desirable (Tracey, 1997).

2.3.4.1 "360 degree" appraisal

Colleagues (doctors, nurses, receptionists and others) are increasingly being used for 360 degree feedback on doctors (King, 2002). Originating as a management tool, the original axis of 360 degree feedback was vertical in a hierarchical system, allowing superiors and juniors to comment on a manager's performance. In General Practice the axis may be conceived as more horizontal, seeking the perceptions of patients and colleagues. Such systems have been used by some forward looking practices to obtain management information to enable the practice to focus on areas which are perceived as most important and least well performed (Griffin, Sanders *et al.*, 2000).

2.4 Assessments of clinical practice

Assessment of competence and performance is a huge topic, and here I am going to concentrate on the formative and summative assessment of GPs as caregivers and

Evaluating Process and Outcome in the Education of General Practitioners

educators. The Regulatory Body for doctors is the General Medical Council (GMC).

The GMC directs doctors:

You must work with colleagues to monitor and maintain the quality of the care you provide and maintain a high awareness of patient safety. In particular, you must respond constructively to the outcome of reviews, assessments or appraisals of your performance.

Likewise, there are similar expectations of a constructive response in education. In both situations the importance of appraisals and assessments that reflect the realities of practice cannot be overemphasised.

Modern clinical practice is too complex for single global ratings to be meaningful (Streiner, 1985). McKinley *et al* (2001) make the point that assessment benefits from being multifaceted, with performance review and selected competence sampling. They discuss the flaws inherent in over-dependence on subjective opinions (Baker, 1997), and acknowledge that ‘...identification of poor practice through monitoring of routine data may be insensitive and inconsistent’ (Frankel, Sterne *et al.*, 2000).

Believing that

..the cornerstone of medical practice is the consultation . . . as all else in the practice of medicine derives from it.

They advocate that the monitoring of clinicians should focus predominantly on the direct assessment of consultation performance by peer observers (McKinley, Fraser *et al.*, 2001).

Perceptions, especially self-perceptions in the form of confidence ratings, have an acknowledged place in appraisal and assessment of the clinician (Newble, Jolly *et al.*, 1994). There is considerable experience with assessment of vocational trainees’ competence using rating scales, such as the new Manchester rating scales for vocational training in general practice, but the experience suggests that at least 23 main rating scales are needed to achieve systematic assessment of vocational trainees (Difford and Hughes, 1991).

The GMC is the final arbiter of “quality” and makes assessments about training processes and about doctors, including any of those that I refer to as the “Trainer” or “Trained Registrar”. (For the last thirty years, most UK GPs have been vocationally

Evaluating Process and Outcome in the Education of General Practitioners

trained, so “Trained Registrar” could refer to most GPs in practice, but I am most interested in those who have been trained in the last 10 years, as over time the effects of training diminish in relation to the influence of experience (Van Leeuwen, Mol *et al.*, 1995).

2.4.1 Attending to underperformance

A disproportionate amount of the GMC’s workload has up to now concerned the detection, performance management and disciplining of under-performing, negligent and professionally unacceptable doctors. Most of the information, which leads the GMC to deem that there is a case for a doctor “to answer” at a Disciplinary Hearing, arrives indirectly via the perceptions and assessments of others. Except for those few cases, which end in Committee hearings, the GMC has little by way of direct relationship with individual doctors.

2.4.2 Appraisal and Revalidation for all doctors

Recently, however, the GMC has introduced a programme of annual appraisals for all doctors (DoH, 2002), paving the way for 5-yearly revalidation (GMC and DoH, 2002). Revalidation is long-established in North America (Bashook and Parboosingh, 1998), where the clinician has to produce evidence of performance. It is believed that in the UK, by contrast (Southgate and Dauphinee, 1998), revalidation will involve a form of 360 degree appraisal, but this has yet to be confirmed.

GMC takes an active interest in the quality of trained Registrars, entering their names on the specialist register, and where there is evidence for concern about a young doctor’s performance in the early years after training, the GMC will ask questions of the Trainer about the training process.

2.4.3 Assessments of educational practice

Regulated by JCPTGP, local PGMDE departments are the regulatory bodies for GP Registrar training, and are responsible for appointing and re-accrediting GP Trainers. PGMDE is principally interested in the quality of training offered by the Trainer. Two important pieces of evidence are the performance of the Registrar in the

Evaluating Process and Outcome in the Education of General Practitioners

nationally obligatory summative assessment, and the Registrar's perceptions of the training.

The quality of each GP Trainer, performing as an educator and as a clinician, is assessed by PGMDE, who rely in part on the direct evidence of a visiting team of peers, but this team in turn relies heavily on the evidence (perceptions) of colleagues and learners and to a lesser extent on the evidence of patients. In the future, as Primary Care Trusts, (PCTs) take more responsibility for clinical appraisal, and for quality assessments as part of their clinical governance responsibilities, many PGMDE teams intend to rely on PCT evidence for clinical performance and to concentrate on educational performance.

There is a reciprocal relationship between GMC and PGMDE, sharing concerns and relying on each other for assessment of doctors' performance in patient care and education respectively.

2.4.4 Evaluations of performance across the medical educator's roles.

In these first two chapters we have seen how the roles of patient care and teaching draw on some of the same qualities, and I have proposed that similar quality frameworks can apply. Bush (1983) suggests 'role' is produced by interaction between 'position' and 'function' and the behaviour and constructs of the individual. Similarly, not everyone will have the same expectations of the role incumbent.

The complexity of human interaction is further emphasised by Hodgkinson and Issitt, (1995) criticising concepts of professionalism, which draw on defined competencies. They emphasise the role of experience in constructing our view of the world and suggest that knowledge and understanding do not merely 'underpin' performance but are in a complex iterative relationship. Competence varies, according to context.

The GP Trainer is evaluated in care-giving and educational roles. Accountability is demanded in both roles, but public demand imposes more criterion-based summative assessment on the physician role. By contrast, education, being a formative discipline, emphasises formative assessment in development. Assessments lean on self-evaluation, but need to be triangulated and repeated over time, as competence is contextual, and is a crucial component of the all-important performance.

**CHAPTER 3
PHASE ONE RESEARCH**

Constructing a Framework for Educational Behaviours

The first research question was, “What Educational Behaviours on the part of GP Trainers have lasting value in the formation of GPs?”

3.1 Conceptual: Phase One Design

The aim of the first phase of my work is to construct a framework, whereby to categorise helpful and less helpful aspects of the learning culture, from a long-term perspective.

3.1.1 Design problems for categorising process

One-to-one teaching is more difficult to categorise than classroom teaching, as there is more homogeneity about groups of learners than individuals. Also, the one-to-one process of teaching of Registrars by Trainers is not a ‘stand-alone’ educational process; learners are simultaneously accessing day-release training in VTS groups.

Bligh’s (1992) work on GP vocational training suggested 3 principal factors influenced GP trainees readiness to learn: enjoyment and enthusiasm for learning; a positive self-concept as a learner; and a reproducing orientation to learning.

A further consideration for research design is that the one-to-one learning relationship between GP Registrar and Trainer lasts for up to one year; a lot can change in that time if the relationship evolves rather than remains static.

Any curriculum of general practice is so broad as to be undeliverable as an entity. This excludes any process analysis that is tied to curriculum delivery, as each Trainer and Registrar will select different parts of any notional curriculum. Trainers are encouraged to develop needs-based teaching, (JCPTGP, 2003) and to be innovative and imaginative about adapting their outline curriculum plans on a week-by-week basis and so there is a huge diversity of approaches.

3.1.2 Possible approaches

Three possible approaches occurred to me. The first option would have been to adapt an already validated framework from a parallel field of research. If a suitable framework exists, then there is much to be said for demonstrating the applicability to this situation, making any necessary adaptations, and revalidating. The second option was for me to draw on my experience of General Practice education, and that of colleagues; to survey the GP literature for factors that are believed to be important; and to construct a framework based on commonly held ideas and theory. The third option, which was the one ultimately selected, was to ground new theory in the experiences of learners in General Practice. I list here the pros and cons as I saw them.

Table 3.1: Advantages and disadvantages of importing and adapting a framework

Advantages of this approach	Disadvantages of this approach
Could find already validated method, which would have credibility	Needs to be truly relevant in the unique situation of GP training
Very time-effective to adapt rather than invent	Adaptation often invalidates previous validation so new process needed
	No suitable framework found!

Much is written about professional knowledge and competence (Eraut, 1994), and I searched the Medical Education literature, and that of other Health Care Professions, including Nursing, Professions Allied to Medicine, and Social Work as well as practitioners in the teaching professions in Higher Education and Further Education, to look for frameworks that could be adapted for my purpose. My searches included the following databases: Medline, Cinnahl, Embase, TimeLit, ERIC, Psychlit and Sociofile. Search terms included Professional Development (CPD); Continuing Medical Education (CME); Higher Professional Education (HPE); Learning Styles; Vocational Training for General Practitioners. I also used search terms such as General Professional Education, Workplace Education, Apprenticeships, One-to-one Teaching, and Music Education (this last because it happened to offer a parallel of independent practitioners being educated in a one-to-one setting).

Evaluating Process and Outcome in the Education of General Practitioners

A Delphi study carried out with GP Trainers, Registrars, and non-training principals, looked for four key attributes of GP Trainers (Munro, Hornung *et al.*, 1998). The very attribute in which I am interested, teaching quality, is a component of their framework, but is not sufficiently sub-categorised for my purposes. The other key attributes are Interpersonal relationships; Professional development; and Personality.

Beyond the world of General Practice, I did not find a framework that seemed fit for my purpose. Buchler’s (1961) framework of three domains of Technology, Craft and Art offered some promise.

Table 3.2: Buchler’s Typology of Method (Buchler, 1961)

Whilst this is a helpful division for GP Trainers to consider, it would need another axis to construct a usable matrix framework, and as such, it offered little in research terms over and above designing a fresh framework.

Joyce and Showers’ (1980) constructed a 5-part model of a craft orientated approach.

**Table 3.3: Craft Orientated Approach to Competence in Education
(Joyce and Showers, 1980)**

The logical approach was to test a matrix constructed out of Buchler’s typology against Joyce and Showers’ craft orientated approach.

Table 3.4: Matrix of Method against craft orientated approach to competence

	TECHNOLOGY	CRAFT	ART
Presentation of theory or description of skill or strategy			
Modelling or demonstration of skills or models of teaching			
Practice in simulated settings			
Structured and open-ended feedback			
Coaching for application			

This hybrid was the best adaptation of existing frameworks that I could fashion for my purpose, and I am trialling adaptation in a teaching-the-teachers course. However, it remains a hybrid, having neither the advantages of being tried and tested in the field nor those of being specifically designed for my purpose, and ultimately I decided that it could not measure up against the latter.

I began to look at broader and vaguer conceptual frameworks and turned to the well-known work by Bernstein (1971), “On the classification and framing of educational knowledge”. Here he talks about the concepts of *classification*, (strong, where there are strong boundaries between different contents of the curriculum; weak, where the boundaries are weak), and *frames* which relate in similar fashion to the strength of the boundaries between what may be transmitted and what may not be transmitted between teacher and pupil.

I tried in vain to apply this theoretical framework to the analysis of process in General Practice education. The concept of *frames* was helpful, but the looseness of curriculum in General Practice learning at this level, served to undermine the use of Bernstein’s framework here. I was interested recently to discover that others have managed to apply Bernstein’s concepts in quite diverse areas of educational research (Walford, 2001).

Evaluating Process and Outcome in the Education of General Practitioners

Another older framework with much to commend it for research on educational process is that found in Parker and Rubin’s (1966) “Process as content: Curriculum design and the application of knowledge”. They distinguish:

- 1. Processes that expose the student to a particular body of knowledge:
formulating questions, reading, observing, listening, collecting evidence, discovering principles.
- 2. Processes that allow the student to extract meaning from the body of knowledge:
analysing, experimenting, reorganizing, consolidating, integrating.
- 3. Processes that enable the learner to affix significance to the knowledge:
inferring generalization, reconstructing, relating to other situations, testing for usability.

The implicit hierarchy of process in the Parker and Rubin framework had potential for my work, but crude experimentation with categorising some of my own teaching showed how difficult it would be to base research on such divisions. Many educational interactions involve all three processes simultaneously.

I was interested in the possible application of the Dreyfus and Dreyfus (1986) model of levels of proficiency:

Table 3.5: Dreyfus & Dreyfus levels of proficiency

Stage 1	Novice
Stage 2	Advanced Beginner
Stage 3	Competent
Stage 4	Proficient
Stage 5	Expert

Attracted to the way that Patricia Benner (1984) had applied this framework in her research on how nurses learn, “From novice to expert”, I reflected on her analysis of the domains of nursing practice, pondering if these could be adapted to studying General Practice:

Evaluating Process and Outcome In the Education of General Practitioners

Table 3.6: Benner’s Domains of Nursing Practice (Benner, 1984)

The Helping Role
The Teaching-Coaching Function
The Diagnostic and Patient-Monitoring Function
Effective Management of Rapidly Changing Situations
Administering and Monitoring Therapeutic Interventions and Regimens
Monitoring and Ensuring the Quality of Health Care Practices
Organizational and Work-Role Competencies

Reluctantly, I decided that the framework leant itself more to personally observed case-study than the work which was here planned.

Another framework that is in some ways particularly apt is that of Gerald Grow (1991), looking at the ways in which tutors adapt to learner stages.

Table 3.7: Matching Learner Styles to Teacher Stages (Grow, 1991)

This very useful framework homes in on a particularly important attribute of the General Practice educator, but I rejected it as the basis for research, on the grounds that it is too specific to cover all the generalities of the situation of General Practice training.

In the medical Education literature, the most promising framework came from ‘How Doctors Learn’ (Slotnick, 1999), but this framework relates more to the individual process of learning and less to the role of the Trainer in promoting that process of framing questions and seeking answers.

Evaluating Process and Outcome in the Education of General Practitioners

I rejected as too reductionist for my purpose the medical curriculum frameworks, such as those of ‘Professional Development’ (While and Attwood, 2000), or confidence rating scales (Difford and Hughes, 1991).

Having established that there was no suitable framework to evaluate the learning culture in General Practice, there was a need to construct one.

Table 3.8: Advantages and disadvantages of designing a framework conceptually

Advantages of this approach	Disadvantages of this approach
Relatively easy to construct a framework based on literature, personal experience and views of colleagues	Lack of originality and freshness in proceeding down well-worn paths
Has the face validity of consensus in educational practice	Unlikely to be consensus amongst educators on the important items to test
Likely to cover all the items that fellow medical educationalists deem to be relevant.	Likely to miss any relevant aspect of situated educational practice that had not previously received attention.
Allows me to ride my hobby-horses!	My personal subjectivity likely to introduce considerable bias into the approaches selected

The literature of General Practice teaching is rich in possible frameworks on the tasks of the GP Trainer (Erasmus, Coetzer *et al.*, 1977; Bligh and Slade, 1996; Snadden and Thomas, 1998; Spencer and Jordan, 1999; Munro, 1998; Ker and Snadden. 2001).

There is also much else that is highly relevant in the literature of Medical Education (Irby, 1990; Shapiro and Talbott, 1991; Davis, Thomson *et al.*, 1992; Davis, Thomson *et al.*, 1995; Barrington and Silagy, 1996; Bashook and Parboosingh, 1998; Ram, Grol *et al.*, 1998; Boaden and Bligh, 1999; Black and Macdonald, 2000).

Working on the idea of constructing a framework that addressed issues of learner-centred teaching, I constructed a template based on one which had earlier been designed in the Aylesbury Trainer Group, mapping the dimensions of the ‘priority objectives’ framework and updating them in terms of preparation for Good Medical practice. I abandoned the exercise when the matrix reached 8 columns and over 40 rows! Reductionist, and far too cumbersome for use in practice, this sort of framework is much better at trapping content than process. It does however have the merit of showing whether some of the less easy areas of the curriculum are taught:

Evaluating Process and Outcome in the Education of General Practitioners

for example, it is easy to see where the Trainer has been working on ‘Personal and Professional Development’.

After much deliberation I decided the way ahead lay in constructing a framework based on the experiences of former Registrars in training.

Table 3.9: Advantages and disadvantages of starting afresh; grounding theory on research findings

Advantages of this approach	Disadvantages of this approach
If done properly, the relevance of the framework to practice is assured by a process which is grounded on learners’ experience	Subjectivity is inherent in the process of coding, however rigorous the technique
This approach should pick up any previously unrecognised aspects of educational process which are important	It may not align comfortably with commonly held views and practice
Product of this process has an intrinsic ‘freshness’ as it derives from learners’ experience	It is very time-consuming and laborious

The chosen method was to conduct semi-structured telephone interviews of past GP Registrars of one district to determine perceptions about what aspects of the year in a training practice have made a lasting contribution to subsequent performance as a GP. Using grounded theory methods, emergent themes around trainers’ behaviours, perceived as more or less helpful by the learners, led to the construction of categories and dimensions of educational behaviours.

3.1.3 Choice of Research Subjects

Aylesbury VTS Registrars were chosen as a convenience sample because the local Trainer group there had spawned the work. The Trainers were able to provide contact details for former Registrars, as nearly all kept in touch since training. Research subjects knew me or knew of me, so it was easier for me to get participation from subjects at different stages of independent practice as trained doctors. I wanted homogeneity of VTS day release scheme, so that I was picking up differences between Trainer input not VTS input into training.

There are obvious disadvantages to this approach, in that there may be problems of generalising findings from one locality, and there is possibly increased potential for bias when the researcher is known to the subjects. For this reason, I did not approach any of my 'own' former Registrars, with the exception of Dr. Graham Easton, who helped me pilot the questionnaire (Appendix 14) and helped with later stages of the research.

3.1.4 Choice of Interview Method

Telephone interviews were chosen for the convenience of research subjects. It is very difficult to get busy GPs to participate in demanding research, but they liked the fact that I was able to offer them a 20minute phone call at any time of their convenience.

3.1.5 Choice of Methodology

Grounded theory was the obvious choice of methodology here for the reason that pre-chosen structures had been rejected in favour of trying to establish what structure emerged from interview data. The decision to get a second researcher to interrogate the data independently for coding purpose, was taken in the belief that a somewhat sceptical medical audience, many of whom are unfamiliar with qualitative research, would have more faith in the interpretation if two researchers had worked independently. Reproducibility is not a *sine-qua-non* of the paradigm of Grounded Theory (Strauss and Corbin, 1998).

3.2 Measurement: Phase One Methods

Interview transcripts were submitted to *open coding* ("the analytic process through which concepts are identified and their properties and dimensions are discovered in the data."). *Categories* ("building blocks of theory that stand for the central ideas in the data") began to emerge, and their properties became definable. The interviews also demonstrated the *dimensions* ("the range along which general properties of a category vary, giving specification to a category and variation to a theory.") (Strauss and Corbin, 1998).

I designed and pilot-tested a semi-structured interview (Appendix 14), working with a former Registrar, Graham Easton. Trainers provided contact details for former Registrars. I invited all of these doctors to be interviewed over the telephone for 20

Evaluating Process and Outcome in the Education of General Practitioners

minutes at the time of their choosing. Interviews were recorded, transcribed and analysed.

I worked manually on printed transcripts, immersing myself in the data for several weeks, following a process of questioning the data and theoretical sampling, proceeding to open coding, and subsequently axial coding to integrate categories with subcategories. When it appeared that theoretical saturation had occurred, (defined by Strauss and Corbin as “the point in category development at which no new properties, dimensions, or relationships emerge during the analysis”), I conducted six further semi-structured interviews with selective coding to affirm the eight selected 'categories' and their dimensions.

We tested these categories and dimensions for face validity by discussion in different forums. Graham Easton made a presentation to the Trainers and some current Registrars (including research participants) at the VTS annual general meeting, and the work was presented at 5 other regional, national and international meetings of medical educators. All categories were meaningful to those involved with vocational training in General Practice (Peile, 2001).

A sample of 10 transcripts was then examined independently and in depth by Dr Neil Johnson (who was blind to the categories I had found), using the same process of open coding. Discussion between the two coders revealed that there was close agreement on the important categories and the only differences were semantic, and mainly concerned a single category, where for myself the central point was personal and professional development, and for Neil Johnson it was the influence of the Trainer modelling behaviours. It proved easy to align these emphases into a single category.

The categories are summarised in Table 3.10, which for ease of reference is also reproduced as Appendix 3.

Table 3.10: Categories and Dimensions of Educational Behaviours of lasting value to learners (Peile, Easton *et al.*, 2000)

CATEGORY		<- DIMENSION ->	
1	Training or Education	Problem-Based Approach Teaching based on approaches to problems which are not limited to present-day contexts	Emphasis on Managing Disease Teaching focused on current policies for disease management
2	Style Spectrum	Wide variety of styles Learner exposed to different consulting styles and role-models in tutorials	Narrow range of styles Teaching dominated by personal style and behaviour of Trainer
3	Space for Reflection	Encouraging reflective practitioner Safe environment to learn from mistakes	Protocol driven behaviour Black and white approach adopted where learner is expected to adhere to guidelines and elements of blame culture likely.
4	Modelling Personal Development and Team Skills	Personal development and team management skills taught Guided learning of skills like time management, assertiveness, boundary-setting	No emphasis on team behaviours Little attempt is made to help learner understand the importance of team-working and the areas of personal development that are involved
5	Learning Cycles	Learning cycles completed A culture exists in the practice where reflection, audit, assessment all promote change and re-evaluation	Haphazard change Culture is reactive to external pressures, and little evidence of information about the practice inspiring meaningful change
6	Family practice in context	Contextualised Learning Trainer introduces the broader dimensions of family and health expectations	Emphasis on presenting problem Focus remains on sorting and shifting
7	Flexibility	Learner centred approach Trainer listens to trainee and positively seeks out their educational needs adapting the training accordingly	Trainer centred approach Trainer adopts rigid structure with fixed views on the educational diet to feed trainees
8	Feedback	Sensitive feedback Both positive and negative feedback delivered where appropriate, stimulating confidence in the learner, and encouraging change	Inappropriate criticism Feedback either inadequate or misplaced or poorly delivered, often not timely or specific enough to be useful to learner

A full description of these categories, with some quotes from the interviews is to be found in (Peile, Easton *et al.*, 2000). Here I will try to distil the essence, as understanding the categories and dimensions is crucial to understanding the work that follows.

3.2.1 Category 1: Training or Education

In the early days of their work, Registrars wanted some factual teaching to help them find their feet in General Practice, but very soon they found it more valuable to learn approaches to problem solving and generic approaches to managing acute and chronic disease in General Practice. The most valued educational approaches were those of using problem cases to evolve generalisable strategies and approaches. This necessitates a Socratic style of teaching.

3.2.2 Category 2: Style Spectrum

The former Registrars were unanimous in valuing the wider variety of styles. A culture valuing difference in the practice helped underline the learning that there is not a single 'right way' of doctoring; a misconception sometimes arising from previous didactic hospital teaching. Two processes were particularly highlighted as useful: joint consulting sessions with all the doctors in the practice throughout the year, (not just 'sitting in' at the beginning), and working alongside non-doctors, as well as doctors consulting, helped to define for trainees the contribution that doctors can make.

Crucial here for Registrars was the understanding that there are many different styles of practice, each with its merits and drawbacks, and that it is permissible to experiment with different styles during the training year, and essential to appreciate that styles of co-workers may differ and yet be valid.

3.2.3 Category 3: Space for Reflection

Challenge is very different from blame and challenge within a safe environment is essential to nurturing reflection, and welcomed by learners.

Evaluating Process and Outcome in the Education of General Practitioners

There is a rather uncomfortable paradox here about protocol driven behaviour. It is a generally accepted aspect of 'good practice' that the doctor conforms to accepted protocols and guidelines in his or her behaviour (Wensing, van der Weijden *et al.*, 1998) and yet, our research suggests the most helpful behaviour in training practices is encouragement for the learner to reflect rather than immediate reference to protocols and guidelines. Later, it may be helpful to refer to guidelines but they can have a stifling effect on reflection if they are cited as the primary response to questions brought up by the learner. Competence, as we determined earlier is a complex interaction of knowledge, understanding, and reflecting on experience: here we see learners endorsing the idea that professional competence cannot be formulaic.

3.2.4 Category 4: Modelling Personal Development and Teams

The illustration in (Peile, Easton *et al.*, 2000) of personal development impacting on learning to work in teams is the way one doctor chooses to deal with telephone consultations. This affects the other partners and the reception staff, as well as the patients. In learning to adapt his or her behaviours to take account of the needs of others and at the same time to be aware of the importance of self-preservation, the doctor's personal development has to include team skills.

Personal skills like appropriate assertiveness and boundary setting as well as the team skills crucial to successful change management, can all be modelled by Trainers, and the interviews yielded numerous examples of this happening. As important as the modelling is the skill of making explicit the learning material which might otherwise pass unnoticed. The value that learners derived from team meetings or partnership discussions seemed to come not merely from attending, but from discussing the issues with Trainers.

3.2.5 Category 5: Learning Cycles

Interviews with the former Registrars endorsed the lasting value that is placed on an evidence-based culture where audit and assessment help to complete learning cycles. Many commented on the difficulty of completing audit cycles within a year, a problem that has been subsequently addressed in changing the regulations for

Evaluating Process and Outcome in the Education of General Practitioners

summative assessment to ensure that candidates assess the effect of the changes in practice, made as a result of audit findings (McKay, Lough *et al.*, 2002).

It seemed from the interviews that audit projects were a metaphor for learning: Registrars really valued a cyclical approach to learning which involved some degree of experimentation and re-evaluation. Proactivity seemed to be one antidote to low morale in the changing culture of General Practice, and learning to change methodically on an audit base was preferable to a constant feeling of 'being changed'.

3.2.6 Category 6: Family practice in context

Registrars have selected family practice often on account of an interest in the broader family and social dimensions of medicine. They valued Trainers teaching them about the context of family medicine. This was an area where the whole teaching practice could contribute, with GPs, nurses and receptionists all helping learners to understand the broader dimension to the presenting problem.

3.2.7 Category 7: Control and Direction

There are some distinctions to be drawn here between learner-centred education and learner-directed education. Good education can be learner-centred whoever directs it, but learners valued an increasing role in directing their education as the year progressed, whatever their starting point on the scale of dependence to independence as a learner. By contrast, there were isolated descriptions of Trainers abdicating responsibility for the training process, so that the learner felt unguided.

3.2.8 Category 8: Feedback

This category was almost omitted on the grounds that it is rather stating the obvious. But to have left out feedback would have meant being unfaithful to the principle of grounded theory, namely to seek out and report what the research data was saying. Time and again, interviewees emphasised the importance of good feedback.

There was nothing new in what they had to say about effective feedback. They wanted frequent feedback that is timely, specific, and constructive. Learners who felt they had an honest triangulated picture of their performance appeared to be more comfortable in their learning.

3.3 Reflection: Pulling the Phase One threads together

Since they were published some two years ago, these categories have been widely debated, at least in the Oxford Region, where they are on the curriculum for the new Registrars' introductory course, and the new Trainers' course. The learners pick up on the need to progress as rapidly as possible to learning about approaches rather than policies (S Plint, 2002, personal communication). The teachers focus more on how to help learners to become reflective (R Flew, 2002, personal communication).

Much of the discussion around the first phase of my work has centred on reflectivity. It seems that reflectivity is central to the teacher's art in Medicine (Brigley, 2002), as in other forms of education (Roth, 1989). It is for this reason that I wanted to adopt or develop an instrument to look at reflectivity in the second phase of this work. Reflection-in-action and reflection-on-action are habits, which are easier to demonstrate than describe (Schön, 1987).

I intend later to test the hypothesis that the reflective GP Trainer, by modelling his or her reflecting behaviours, induces reflective practice in the learner.

It is crucial that Trainers are aware of the power of their personal modelling. It is not necessary for Trainers to be paragons of excellence to model effectively; what is needed is for them to be aware of the effects of their attitudes and behaviours on the Registrars, and to demonstrate the ways in which they are addressing their own personal and professional development within the practice context.

3.3.1 Relating the Phase One research findings to other General Practice Literature

The responses in the interviews we conducted were mostly positive for the factors identified by Bligh (1992) and suggest that those doctors we interviewed were potentially receptive learners. One quote from a GP Registrar illustrates this:

After 8 years of doing hospital jobs then I had one year in practice and my whole approach to people was completely changed by my year in practice. And it was quite incredible really, I feel it certainly changed my whole approach to consultation and everything.

Evaluating Process and Outcome in the Education of General Practitioners

It would seem that Category 1 (Training or Education) alongside Category 4 (Modelling personal development and team-working), are of prime importance in respect of the attributes researched by Munro *et al.* (1998), but it is likely that Category 7 (Flexibility) and Category 8 (Feedback) are also highly relevant.

At the time this research was conducted, the learners' need for more attention to completing the learning cycle often reflected their experience of incomplete audit cycles. The audit requirements for summative assessment have now changed, and it is to be hoped that learners will benefit from implementing change and re-auditing. This will require new skills on the part of Trainers, who have not always been very adept at guiding and marking audit projects (Lough and Murray, 1997). The hope is that as project work becomes cyclical, so learning in general will follow the process of action, evaluation, revision, re-evaluation.

3.4 Summary

In order to research effectiveness of different processes of education by GP Trainers, it was necessary to have a framework within which to analyse these processes. After reviewing the literature, and finding no suitable frameworks to adopt or adapt, I constructed a framework, based on delayed Interpersonal Recall of former GP Registrars, using a grounded theoretical approach. This framework has content and construct validity, and though somewhat complex for trainers and assessors to assimilate, it has proved robust in use throughout Oxford Deanery.

CHAPTER 4
PHASE TWO RESEARCH
Assessing Educational Behaviours in Training Practices

4.1 Conceptual: Phase Two Design

The second research question was, “How can prevailing Educational Behaviours on the part of GP Trainers be assessed in training practices?”

4.1.1 Sequential stages of Phase Two

The objective of Phase Two is to identify how different characteristics of education may be reflected in assessable dimensions of training practices. There are four parts to this phase:

Phase IIa (Understanding how judgements about teaching are made on the basis of inspection reports) An initial study to understand the process of reporting and accrediting training practices.

Phase IIb (Pilot) A small-scale in-depth study of a selection of training practices in the Oxford Deanery.

Phase IIc (Roll-out) Having trained all current Team-leaders in the methods refined in Phase II a, all training practices accredited and re-accredited in Oxford Region over the course of 18 months are assessed on the selected categories of training behaviours.

Phase IId (Refinement) As a result of the Phase IIb experience, further steps are designed to research and implement the assessment of educational behaviours in training practices.

4.1.2 Limitations imposed by research context

Pragmatic considerations again influenced the design of this stage. It was apparent right from the beginning, that in these highly-stressed times in General Practice, the amount of time that could be invested by training practices in research co-operation was very limited, and this assumption was amply confirmed later.

Evaluating Process and Outcome in the Education of General Practitioners

Options for conducting assessments of training in practices are outlined in Table 4.1. Before I had weighed up the options from the research viewpoint, the matter was decided, as the Director of General Practice Training took an executive decision to incorporate the research assessment into the routine visits. The thinking behind this decision underlines one of the particular aspects of this research: the process is aimed not only at increasing knowledge, and underpinning scientific assessment of educators, but it has a clear educational objective in getting Trainers to think more about educational process.

4.1.3 Compromises in the research endeavour

There is a trade-off in the research design: keeping the research ‘pure’ and uncontaminated limits the dissemination of the framework to look at educational process, and prevents the body of Trainers familiarising themselves with this framework in the course of their duties on peer assessment visits. On the other hand, allowing Trainers to become thoroughly familiar with the process categories before they are properly validated by research, risks significant contamination getting in the way of validation. My experience of Action Research (Peile, 2000) has been influential in allowing me to see the value of the research endeavour proceeding iteratively with improving practice.

4.1.4 Accrediting the Trainers locally: Personal experience of the researcher

As a former GP Trainer, I had been a member of peer review teams since 1986, and in 1996 I became a Team-leader. My discomfort with the emphasis in peer review visits on *structure* of training rather than *process* of education was the stimulus to the work described in Chapter Three. When I started to think about the whole accreditation process, I began to ask questions about the way in which Team-leaders reports are evaluated, and how the Trainer Selection Committee makes an assessment for accreditation or re-accreditation of a training practice. I discussed these thoughts with the then Director of Postgraduate General Practice Education, Neil Johnson, and together we decided on further research.

Table 4.1: Options for research methods Phase Two

Possible Method	Advantages	Disadvantages
<p>1</p> <p>Researcher conducting special visits to all practices</p>	<ul style="list-style-type: none"> • consistent method • high quality paperwork • direct comparison between sites possible • could keep research categories unpublished and undisseminated . 	<ul style="list-style-type: none"> • less easy to generalise on one observer's reports • less dissemination of categories in the research process • extra demands on practices likely to lead to poor co-operation
<p>2</p> <p>Large training exercise whereby all Trainers trained to conduct special peer visits to assess educational process as part of the research exercise.</p>	<ul style="list-style-type: none"> • all Trainers become rapidly familiar with new categories - helps Trainer formatively. • wide experience allows for many views on usefulness of research categories • large numbers of assessments assured 	<ul style="list-style-type: none"> • very expensive to train and operate as a research exercise • very time consuming for Trainers • likely to be poor quality paperwork and data • too many observers impairs the consistency of research method
<p>3</p> <p>Training Team-leaders so that research elements can be incorporated into routine assessment visits</p>	<ul style="list-style-type: none"> • allows graded introduction of research concepts into practice with space for adaptation • nonetheless, Trainers get used to the new framework gradually in the course of peer visiting • educational process assessment incorporated into the routine visit is mirroring the changes I hope will happen in practice • constitutes a form of 'action research' with practitioner involvement as researchers 	<ul style="list-style-type: none"> • needs effective training for the 12 Team-leaders • even with Team-leaders trained, the other members of the team may be lost if not familiar with the evidence they are collecting • scope for poor paperwork and data collection when team leaders are trying to cram research exercise into already crowded visit schedule • ethical concerns about research being incorporated into compulsory routine visits. • prior familiarity with the research framework can contaminate the research, when peers on the visiting team come to be visited themselves.

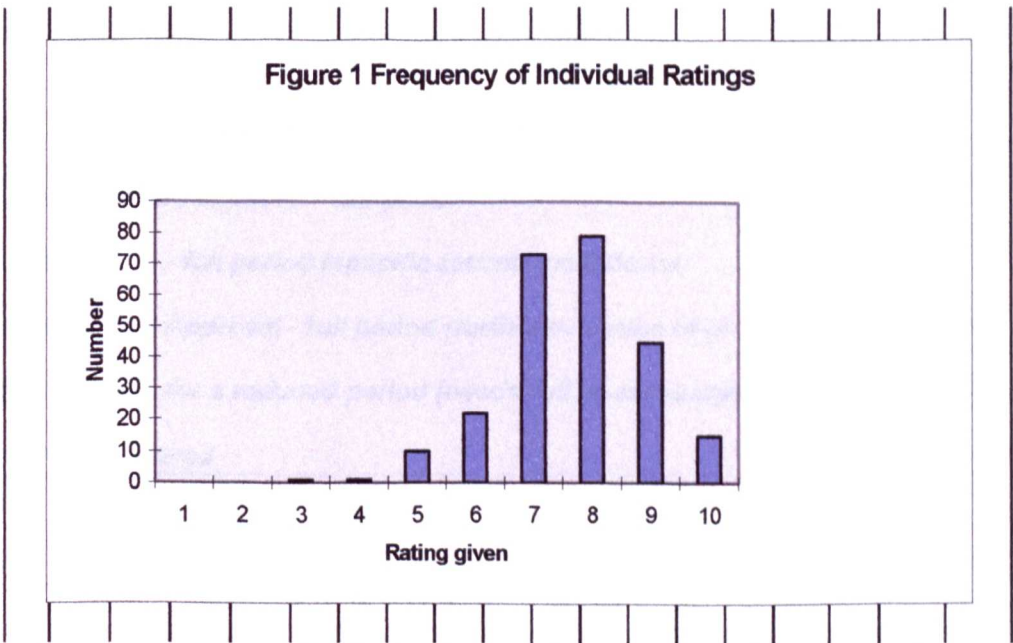
4.2 Measurement: Phase IIa (Understanding how judgements about teaching are made on the basis of inspection reports)

For a period of one year (1999/2000) all Trainer Selection Committee members were asked individually to make a rating, based solely on their reading of the Team-leader’s report, in advance of the committee meeting. We asked for a single global rating on the training a GP Registrar would be likely to get in the training practice, using a scale of 1-10; (1 = dreadful to 10 = exemplary).

The ratings by individual members were not revealed to the Committee, who made their decisions in the normal way. At the end of the year, the ratings by individual committee members were tabulated, for each Trainer considered in committee. A comparison was made with the decisions on accreditation made by the committee.

How do different individuals rate the same Team-leader’s report?

A minimum of four committee members returned ratings for all 46 Trainers who were assessed in committee over one year. The 246 ratings ranged from 3 to 10, in a positively skewed distribution - see Figure 1.



The ratings are tabulated in full in Table 4.3, where the inter-observer and intra-observer variation is apparent.

Looking at the Trainers with the 23 highest mean ratings (= ‘top half’), there is a high level of agreement on their excellence. Out of 123 ratings recorded for this group, all

Evaluating Process and Outcome in the Education of General Practitioners

bar one are between 7-10, most (109) between 8-10. Fifteen of the 16 recorded top scores of 10 points were for 11 Trainers in this group.

At the lower end of the quality ranking (=‘bottom half’), there is also clustering of low rating scores. All bar one of the scores of 6 were awarded to Trainers ranked in the bottom half, and all the recorded scores of 3, 4, or 5 points appear in the bottom quartile.

For each individual Trainer, the range of scores is fairly tight. No Trainer was scored outside a 5-point range at the low end (3-7; 4-8; or 5-9). In the top half the consensus was tighter, with only four out of 23 Trainers being scored across a 4-point range (7-10), and a majority scored within the 2-point range of 8-9 or 7-8. The lower level of consensus at the lower end of the table happens because some isolated scores of 9 points appear scattered down as far as the Trainer ranked 40=, suggesting that there are qualities which are rated higher by some individuals than others.

4.2.1 Comparison with accreditation approval outcomes

There are five possible outcomes of committee deliberation in terms of decisions on Trainer accreditation and re-accreditation, see Table 4.2.

Table 4.2: Key to Committee Decision Outcomes

<i>A = Unqualified approval - full period</i>
<i>B = Approval - full period (specific recommendations)</i>
<i>C = Qualified Approval - full period (further evidence of progress needed later)</i>
<i>D = Approval for a reduced period (needs full re-assessment earlier than normal)</i>
<i>E = Not approved</i>

Evaluating Process and Outcome in the Education of General Practitioners

Table 4.4: Different observers rating team-leaders reports

Table 4.4. Different observers rating team leaders reports
(Rating on a scale of 0-10: each rating value colour-coded differently for ease of visual overview)

(Rating on a scale of 0-10, each rating name colour - context) for each of 10 items. Row 47 = total value of ratings by individual observer. Row 48 = number of ratings accorded by the individual. Row 49 = average rating for that individual

[illegible]

Evaluating Process and Outcome in the Education of General Practitioners

In six cases, the ratings on Trainers could not be compared against approval outcomes and these cases were omitted from analysis. The others are compared by approval grading against centiles for individuals' prior ratings, see Table 4.3.

It can be seen that there is a trend towards lower grades in the lower centiles, and the difference between outcomes for the top half vs the bottom half (by prior ratings), is apparent.

Table 4.3: Graded outcomes of committee decisions by prior (individual) ratings

result	Grade A	Grade B	Grade C	Grade D	total
top quartile	10	0	0	0	10
second quartile	5	4	1	0	10
third quartile	1	6	2	1	10
bottom quartile	3	2	3	2	10
total	19	12	6	3	40

This study looks at the way in which individual members of a committee vary in their appraisal of core documentary evidence. A global rating, based on a simple question, is used, because there is evidence from other settings that global judgements may be more valid than a detailed breakdown into component ratings (Keynan, Friedman *et al.*, 1987; Norcini, Diserens *et al.*, 1990; Gray, 1996; Rothman, Blackmore *et al.*, 1997; Regehr, MacRae *et al.*, 1998).

During the course of this year, no Trainer was refused accreditation or re-accreditation, although some withdrew their application at an early stage in the process.

Withdrawals of some Trainers anticipating unfavourable appraisal, may have contributed to the positive skew of ratings seen in Figure 1.

4.3 Measurement: Phase IIb (Pilot)

In order to establish whether there is an effect of these training behaviours upon subsequent performance of the trained Registrar, it is first necessary to establish that training behaviours can be assessed in training practices.

Team-leaders have bi-annual training days at which the conduct of the visits is reviewed. At one of these sessions, there had been an expressed desire to increase emphasis on the process of the educational activity, and our research appeared to resonate with this, so that Team-leaders were keen to include an assessment of training behaviours on the practice visits. There was sufficient enthusiasm amongst the Team-leaders for the Director to decide that for a trial year, all practices being assessed for training accreditation should have training behaviours recorded under the new categories. This work would be undertaken by visiting teams and reported by Team-leaders, and their findings would form a part of this research project. However, this change in emphasis had obvious training implications for the visiting team, and we decided to pilot the changes on some research visits, before attempting to train Team-leaders and Practice Managers in the new procedures.

Dr Tim Huins, recently retired as an Associate Adviser in General Practice at Oxford-PGMDE, volunteered to help with the pilot visits. After making two visits together, in order to calibrate our assessments, we carried out a further six visits working separately.

Subsequently, we designed a training programme for Team-leaders to incorporate our assessments into routine re-accreditation visits. With the consent of a Trainer undergoing assessment for re-accreditation, we asked a visiting team to pilot the incorporation of the new technique into a scheduled visit, which I attended as an observer. Feedback from the visitors and the visited Trainer was used to improve the training programme.

4.3.1 Research participant Trainers

For the initial pilot visits, we asked to visit practices which had been re-accredited in the past year. As these Trainers mostly had their relevant training records, including a video of a tutorial, readily to hand from their recent visit, we reasoned that these practices were the ones where we would cause the least disruption by our pilot visits. Of the twenty practices invited to participate, ten agreed to do so, and eight were visited. They were each paid a small honorarium in respect of professional time taken up in the research endeavour.

4.3.2 Research Visits and Methods of Assessment

The visits were organised at a mutually convenient time, and Trainers set aside 45 minutes for an interview. We wanted to explore six opportunities for assessing training behaviours. Five of these methods are standard components of training re-accreditation visits. They are: Interview with Trainer; Interview with Registrar; Review of video of tutorial; Inspection of training log and programme; and Inspection of guidelines, protocols, and audits. Used together, these methods can test all the educational behaviours in which we are interested - see Table 4.5.

The other method which was tested was a 'Standard Scenario' designed to pose common educational problems, encountered by trainees, in a format which tested the Trainer's style of response to learner's questions. Answer guides were written, describing examples of possible responses designating "Preferred Behaviour" in different categories, and alternative responses indicating a "Less Helpful Behaviour". (See Figure 2). This assessment is similar to the use of standard patient scenarios in the clinical assessment of GPs (Rethans and Saebu, 1997).

Evaluating Process and Outcome in the Education of General Practitioners

Table 4.5 Examples of opportunities for assessing training behaviours on re-accreditation visits to training practices

<u>Training Log and Programme</u>	<i>Category</i>	<u>Video of tutorial</u>	<i>Category</i>
Is the emphasis (after 1 st month) more on managing disease or developing a problem-based approach?	1	Is the emphasis didactic on managing disease or Socratic, developing a problem-based approach?	1
Is there broad exposure of learner to other doctors and team members through year?	2	Is reflective practice modelled and encouraged?	3
What are the opportunities in the programme for personal development learning about self and teams?	4	Are opportunities used for learning about self and teams?	4
Is learning about audit programme into time-table?	5	Is there evidence of family contextualisation?	6
Does log reveal evidence of encouragement towards self-direction?	7	Is there ad-hoc evidence of Trainer adapting to learner stage?	7
Is Trainer feedback recorded and tracked?	8	Is sensitive feedback demonstrated?	8
		<u>Trainer Interview</u>	
<u>Registrar interview</u>		Does Trainer emphasise need to learn disease management protocols or adopt more of a problem-based approach to teaching?	1
Emphasis in teaching on managing disease, or problem-based approach?	1	How is reflective practice encouraged?	2
Good exposure to other doctors and team members all-year?	2	What opportunities has Trainer offered for learning about self and teams?	4
Reflective practice encouraged and facilitated?	3	Is needs assessment modelled as a part of care process?	5
Opportunities found for learning about self and teams?	4	Is needs assessment a part of educational process?	5
Has learner seen tangible change happen as result of previous audit projects in the practice?	5	What evidence can Trainer offer of encouraging appropriate learner self-direction?	7
Does learner understand educational needs assessment?	5	Is Trainer comfortable giving feedback? - cite examples of positive and negative feedback given	8
Do team encourage contextual thinking about family practice?	6		
Is there evidence of Trainer having encouraged appropriate learner self-direction?	7		
		<u>Records of training</u>	
Comfortable feedback received? Can learner give examples of positive and negative feedback received?	8	Does development of learner's project reveal evidence of understanding importance of completing audit cycle?	5
		Do the protocols & guidelines encourage reflective practice or are they didactic?	3

NB Category Numbers refer to the Categories of Training Behaviours defined in Table 3.10

Figure 2 - Example of Standard Scenario

Standard Scenario 4

Your Registrar is nine months into the training year. She catches you after morning surgery with a problem she wants help sorting out. She has just seen a 24 year old lady, who is significantly unhappy about her large breasts, and wanting surgery, (reduction mammoplasty), which she cannot afford privately. The consultation has revealed that the patient's breasts have contributed to the low self-image, which has been evident over past years, and may well have played a part in a depressive episode last year, (which followed a relationship breakdown), as well as contributing to aching shoulders. The Registrar is aware that cosmetic surgery is classed as a 'Low priority procedure' for referral purposes. This means that she would have to make a special case, and she has promised to ring the patient back after discussing the matter with you, the Trainer.

Problem posed - The Registrar, conscious of a responsibility to husband resources, wants to know if she should refer the patient to secondary care.

Try to describe in detail how you would handle this question, and where it might lead.

.....
Answer Guide

PREFERRED RESPONSES

1. *Asking the Registrar to elicit the pros and cons of agreeing to the patient's request (Space for Reflection)*
2. *Referring to guidelines AFTER learner has had a chance to think out a strategy in her own way- (ie as a reference for checking out how her treatment plan fits in with conventional advice) (Space for Reflection)*
3. *A strategy which involves the Registrar in thinking about the educational issues raised and how she might address them (Control and Direction)*
4. *Constructing an approach to the problem which has wider relevance than just sorting the case in question. (Training or Education)*
5. *Checking that context is considered in responding - rather than producing one answer for all women with large breasts (Family practice in Context.)*

UNHELPFUL BEHAVIOURS

1. *Referring learner straight to authority - eg " Look at the practice protocol" or "Ring up Public Health" (Space for Reflection)*
2. *Didactic advice on what to do (Space for Reflection)*
3. *Limiting Response to this specific case without attempting to guide learning of a generaliseable nature (Training or Education)*
4. *Considering the referral issue in isolation (eg getting hung up on whether reduction mammoplasty is cosmetic surgery) (Family practice in context)*

NB other behaviours such as feedback may be demonstrated in the Trainer's response.

4.3.3 Joint Pilot visits

The first two visits, conducted by both researchers working together, were designed to discover if the format was practical and the method *potentially* reliable.

It was helpful to have one of us observing whilst the other asked questions. Our questioning style in these early interviews was directed specifically towards ascertaining the relevant information about training behaviours, and we observed that we tended to lead our interviews in a manner that shed doubt on the value of the responses. For example, most Trainers would pay lip-service to the value of reflective practice, and so asking about it directly, inevitably elicited “Preferred Responses”. We found that it was possible to ascertain more credible information, simply by conducting the interview in the style of a conventional re-accreditation visit, with which both researchers and research subjects were familiar. Henceforth we adopted the practice of watching for evidence that “dropped out” of conventional interviews, and recording ‘snippets’ which we felt might be indicative of a training behaviour in a particular category.

Formal assessment of inter-observer reliability was impractical, and we relied on a subjective comparison of our field notes to show us if we were picking up on the same information, and weighting it similarly. At the beginning, my colleague was less familiar than I was with the Categories and Dimensions of training behaviours, which I had developed. By the second visit, we found there was close agreement on what constituted relevant evidence and we had calibrated qualitatively how it should be evaluated.

The process of conducting the visit, and feeding back to the Trainer was refined, and we felt confident to proceed to a further set of pilot visits, each working independently.

4.3.4 Solo pilot visits

Modifying the recording sheets simplified the visit process. Being very familiar with the categories, we found that each of us was able to jot down relevant cues, without interrupting the flow in our spontaneous interview.

Evaluating Process and Outcome in the Education of General Practitioners

These visits covered a wide range of training situations, large and small practices, experienced and less-experienced Trainers, and we met learners at different stages of their training. It was this last variable that caused us most concern. We each found an interview with the Registrar in post to be a very valuable source of information, but early learners have very different learning needs from the almost independent practitioner at the end of his/her training year. We needed to contextualise evidence gleaned from interviewing Registrars.

Our research subjects for this pilot phase were selected on the grounds that some had been highly rated at re-accreditation, whereas others had been borderline for approval. The researchers were blind to the recent re-accreditation process, as the names of potential subjects to approach (from the pool of recently re-accredited Trainers) had been supplied by Oxford-PGMDE. One problem that we encountered, however, was that the vast majority of evidence that we recorded was of a positive nature – reflecting “preferred behaviour” rather than “less helpful behaviour”. The data looked rather homogenous, however in all cases we were able to feedback to Trainers not only the positive encouragement for “preferred behaviours”, but also evidence for behaviour in one or more categories that might be perceived as “less helpful”.

Trainers commented that they found the feedback helpful, in particular where it yielded material for reflection on whether it might be helpful to change an established behaviour (see examples in Figure 3 on the right side – fairly weak and very weak).

We found the standard scenarios to be particularly helpful. Trainers seemed to respond in a manner that accorded with their perceived training style. We had confidence in this method and in the video, to reveal evidence of *what the Trainer actually does* rather than *what the Trainer says he/she does*, which comes through at interview. All respondents engaged with their proffered standard scenario in a positive and realistic manner. There is evidence from clinical practice of the reliability of similar scenarios inducing responses from physicians that closely resemble performance in practice (Peabody, Luck *et al.*, 2000).

4.3.5 Designing a Training package

We agreed with a candidate Trainer and an assessment team that we would pilot this work on his re-accreditation visit. The training was designed to make minimal changes to existing visit procedures. We reassured those involved that, apart from the additional Standard Scenario Test, nothing was changed in the conduct of the visit. While asking the same part of the questions as in previous visits, one member of the team recorded evidence of training behaviours.

The visiting team was encouraged to familiarise themselves with the Categories and Dimensions before the visit, and we tried to make the recording forms as user-friendly as possible. A composite form illustrating some examples of evidence was demonstrated (Figure 3). Note that assessors are encouraged to use actual quotes in their snippets of evidence wherever possible.

4.3.6 Incorporating the methods into an actual re-accreditation visit.

Training a visiting team gave us the opportunity to try out a presentation on the Team-leader and written training materials for the other team members. It has to be said that although they had mostly tried to familiarize themselves with categories and dimensions, they struggled to record the data adeptly. However, having done so, they were readily able to enter into the process of weighing up the evidence, and for 6 out of 8 categories their assessment of the Behaviours on a 5-point scale accorded exactly with my own rating (I was present as an observer.) Furthermore, the Trainer in question said that he appreciated the feedback on training behaviours, but the cost of this added value was an extra hour of visitors' time on an already long day.

4.3.7 Training the Trainers

At the next bi-annual training day, much the same training pack was presented to Team-leaders and Practice Managers, in a seminar and workshop. The team who had performed the pilot re-accreditation visit and a Team-leader, who had (as an active Trainer) been assessed on one of the earlier pilot research visits, were able to feed into the discussion.

Evaluating Process and Outcome in the Education of General Practitioners

The outcome of the session was that Team-leaders requested two changes to the visit structure. Firstly, they asked that the grades were reduced from a 5-point grading across the dimensions of a category (very strong; strong; intermediate or no evidence; weak; very weak) to a 3-point grading (mainly preferred behaviour; intermediate or no evidence; mainly less helpful behaviour). This was at the behest of those who had been on the pilot visit, and found a 5-point scale difficult to determine. Secondly, Team-leaders requested that the Standard Scenarios be dropped from the visit schedule, largely because of the workload implications of adding an extra item to a crowded schedule.

With these two changes, the visit format was agreed, and became the protocol for Phase IIc of this research.

4.4 Measurement Phase IIc (Roll-out)

4.4.1 Routine Assessment of training behaviours for a trial period

The Director informed Trainers, who request accreditation or re-accreditation, of the new procedures before they are visited (Appendix 4). The protocol for members of the visiting assessment team is reproduced at Appendix 5.

Early experience of the visits in progress suggested that the format was still proving too complex, as the documentation of evidence for training behaviours was of poor quality. We attempted to remedy this by making the documentation forms more intuitive, in the hope that, as experience of the procedure widened amongst Team-leaders and Team-members, data quality would improve.

Despite our best efforts, the data remained of poor quality for research purposes. Less than 1 in 4 of the first 20 forms returned were usable.

Evaluating Process and Outcome in the Education of General Practitioners

Figure 3: Extract from training pack: composite illustration of evidence recording

Category 1 – Training & Education

Problem-Based Approach <i>Teaching based on approaches to problems which are not limited to present-day contexts (preferred behaviour)</i>		Emphasis on Managing Disease <i>Teaching focused on current policies for disease management (less helpful behaviour)</i>
Very Strong 5 Strong 4 (T)Eating disorders and issues around (V)Prompting for general issues (V)Breast lump mismanagement turned to “How do I deal with things I know nothing about?”	Intermediate/Neutral 3	Fairly weak 2 Very Weak 1 (R)He expects me to know how to manage things from just reading protocols

Category 2 - Style Spectrum

Wide variety of styles <i>Learner exposed to different consulting styles and role-models in tutorials (preferred behaviour)</i>		Narrow range of styles <i>Teaching dominated by personal style and behaviour of Trainer (less helpful behaviour)</i>
Very Strong 5 Strong 4 (R)Everybody’s involved in teaching me and I learn from all	Intermediate/Neutral 3 (L)Not adventurous use of other team members in tutorial plans (SS) No mention of involving other team members	Fairly weak Weak 2 Very 1 (R)He wants me to model myself on him (T)I ask the others to teach on specific subjects (rather than letting them choose)

Key to Evidence sources T = Trainer Interview
 R = Registrar Interview
 V = Video of tutorial
 L = Logs and records of training
 SS = Standard Scenario

Table 4.6: Team-leader's recording on Assessment Visits of dominant educational behaviours for each of the eight Categories (see Appendix 2)

	Category 1	Category 2	Category 3	Category 4	Category 5	Category 6	Category 7	Category 8
Trainers scored as "Preferred Behaviour"	25	17	29	22	16	21	26	24
Trainers scored as "Neutral"	1	3	2	1	0	1	2	5
Trainers scored as "Less Helpful Behaviour"	4	9	0	5	7	3	4	2
More than one entry for a single Trainer!!	5	12	7	9	17	8	6	6
No entry !!	6	0	3	4	1	8	3	4

The forms and instructions were revised and data quality improved to 60% usable forms over the next period. The forms with at least partly usable data on 41 Trainers assessed over one year are tabulated in Table 4.6.

4.4.2 Team-leader's study day

A Team-leader's study day was held to review the project on 8th November 2001, and the minutes of that meeting together with data attachments form Appendix 6. Team-leaders supported the further development of the framework, including suggesting further research to take the project forward, and this is discussed further in Chapter 6.

The significant findings were that most Trainers were found to be showing evidence of preferred behaviours, although all had stronger evidence in some categories rather than others. Team-leaders commented on the difficulty in getting used to the categories and learning to record appropriate evidence but all felt they were improving. All Team-leaders and their teams had significant difficulty in making judgements in some categories. (The pie-charts as attachment 2 to Appendix 6 do not reflect this difficulty adequately, as for the purpose of the study day, imperfect data were included).

There was striking disagreement amongst Team-leaders as to what extent these assessments could be used summatively in the future. I, as the researcher, was at one end of a spectrum of opinion, advocating the development as formative tools for the foreseeable future, but the most experienced member of the group, who had national

responsibilities for assessment process was keen to see rapid development into summative tools.

4.5 Reflection Phase IIa

4.5.1 How consistent are individual judgements?

Attention is drawn to the ratings of the five core committee members, in the centre section of Table 4.3. These five attend more meetings than other committee members, who are only called for individual cases, and so it is possible to look at their internal consistency. The mean scores awarded by these individuals are close, except for the Registrars’ representative, who marked Trainers more generously.

4.5.2 How similar are experienced assessors in their grading?

Looking at the two senior members, the Director scored all Trainers, his deputy 32 of them. Their mean scores were 7.41 and 7.78 respectively. The Director used the range of scores 6-9 almost exclusively (bar a single 5), his deputy was similar except for awarding three scores of 10. Of the 32 Trainers scored by both senior Committee members, 87% were awarded scores by two observers within one point of each other.

Table 4.7: Comparing the two most experienced observers

D= Director DD = Deputy Director

D awards score one point higher than DD	D and DD award same score	DD awards score one point higher than D	DD awards score two points higher than D
4	14	10	4

4.5.3 Assessments and perceptions: discussion in committee – the ‘shadow’ effect on individual perceptions

Committee discussions are intended to be an opportunity for discussants to air their views, to expose them to questioning, and frequently to reformulate them in the light of debate with other committee members. What are the common ways in which individual viewpoints are influenced by committee discussions? Parlett (1991) described practitioners sharing their assessments of particular children:

Evaluating Process and Outcome in the Education of General Practitioners

What was common, in fact, was a tendency for general images of children to become shared among a group of collaborating practitioners. In many cases, a child becomes either favourably or unfavourably regarded. "Halo effects" are further sustained when different professionals hear one another's opinions.

The inevitable reduction of complexity in committee discussions led (Cicourel, 1976) to draw a connection between what happens in discussions of cases and what happens in the transmission of rumours. There is "levelling," with the material becoming "more concise and more easily grasped"; there is "sharpening," meaning that perceptions and reports are increasingly selective; and there is "assimilation," with reports becoming "more coherent" and "consistent with presuppositions." In these ways, Cicourel claims, a complex picture is whittled down to a simpler one, the out-of-character statements are reduced, and the general stereotyped image of the subject is built up and then becomes *the* definition. Some of these concepts are beautifully illustrated in the jury-room discussions of 'Twelve Angry Men' (Rose, 1983).

'Groupthink' is the term (borrowed from George Orwell), which Irving Janis (1982) applied to the decisions often reached by highly cohesive groups. A very tightly knit group can concern itself more with preserving group solidarity than with objectively evaluating all possible alternatives in decision-making. As a result, individual members of the group may subdue any urge to voice dissent.

It is not, therefore, surprising that this research shows that the outcomes of committee decisions cannot be predicted reliably from the uninfluenced view taken by individual members on the Team-leaders' reports. It is also self-evident that there is no point in convening a committee unless it is intended that individuals should modify their views in the light of other opinions – it would be a lot cheaper and easier to send out the Team-leaders' reports and collate individual responses by post or e-mail! One attribute of committee discussion is that it may lead individuals to focus on previously unnoticed details in the report, which then assume enough importance to influence accreditation outcome. This is a likely explanation for some of the discrepancies we found between individual prior ratings, based on a global perception of quality, and outcomes of committee discussion.

4.5.4 Assessments and perceptions: discussion in committee – are all perceptions equally weighted?

Committee members might, perhaps, be expected to defer to the experience of the Director or his deputy, and there are clues in the present study that committee discussion may be particularly influenced by the attitudes of the senior members. Take, for example, Trainer 15. Despite ratings of 8, 9, and 10 points from other committee members, this Trainer received only a qualified accreditation (C), which may reflect the prior rating of only 6 points by the Director. Likewise, the Director had also viewed practice 27 in a worse light than others (rating 6 points), and this Trainer received one of the worst accreditation levels (D) despite a mean score of 7.5 resulting in a ranking of 27th. Sutherland (1994), in his book on Irrationality, picks up on Irvin Janis' concept of 'Groupthink', exemplifying the ways in which leaders can influence committees.

However, the weight of senior opinion influencing the other members of the committee does not explain all cases where prior ratings are a poor predictor of committee outcome. In the case of some other Trainers, notably numbers 23 and 29 (whose accreditation appeared to have been qualified more severely than would have been expected from prior ratings), and 37, 38, and 39 (who received A-Grade unqualified accreditation despite low rankings), the ratings of the two senior members do not appear to have been influential.

4.5.5 Assessments and perceptions: discussion in committee – can we truly exclude prior knowledge from the 'shadow' effect?

As in any Deanery, several members of the Trainer Selection Committee are likely to have at least some prior knowledge of the Trainer or the training practice. Indeed, to date, in the Oxford deanery the local VTS Course Organiser and the LMC representative have been invited to take part in the discussions, in order to ensure that local views are heard. Likewise, the more senior members of the committee are quite likely to have received prior intelligence, as concerns, complaints, or plaudits about any Trainer are most likely to have been voiced by the educational community to an Associate Advisor or to the Director.

Evaluating Process and Outcome in the Education of General Practitioners

Prior knowledge can affect committee debate in two ways; directly and indirectly. The direct route is discouraged in our committee nowadays, as there is tacit agreement that the Team-leader's report should be the focus of debate, rather than any privileged information which might be introduced at this late stage, rendering the proceedings less open and accountable.

Indirectly, any one member, influenced by prior concerns, could pick out nuances in the Team-leader's report, and demonstrate reasons for concern to the other committee members, who might otherwise have missed them. This is an example of looking at the same evidence but in a different light, or through a different filter.

4.5.6 How can we improve the fairness and reliability of committee discussion?

Reflecting on the problem of prior knowledge, I became increasingly concerned about contamination of the decision-making process (see Table 4.8).

Table 4.8: Concerns about 'Prior knowledge' affecting Committee decisions

- Prior knowledge is often vague, prejudicial and unreliable (based on hearsay).
- The introduction in committee of knowledge and views based on personal experience tends to devalue the Team-leader's report: what is the point in going to the trouble of mounting a visit and carefully crafting a consensus report, if it is going to be sidelined by committee member's anecdotes?
- Non-attributable material actually undermines the openness and accountability of the process. Team-leaders' reports are accessible to the Trainer, unlike unminuted committee discussions.
- By the same token, decisions, which are not clearly based on documented evidence, are more open to appeal.

In Oxford Deanery, the local VTS Course Organiser writes a report on the Trainer for the visiting team. This report, to an agreed format, is an open document, which is copied to the Trainer before the visit. Any concerns or issues raised in the Course Organiser's report are then addressed in the course of the visit and may be commented upon in the Team-leader's report.

It would not be difficult to extend this process, and we propose that after the Course Organiser has written his/her report, it should pass up the line for further comment by the responsible Associate Adviser and Director. This then is the opportunity for

Evaluating Process and Outcome in the Education of General Practitioners

senior people to share relevant prior knowledge, or else forever hold their peace. It should not be impossible to table such concerns openly, whilst if necessary protecting the source of privileged information.

The main advantage of concerns being aired **before** the visit, is that there is the opportunity **at the visit** to collect the relevant evidence from interviews, records, and observed behaviour. The committee can then discuss that evidence, rather than nebulous concerns.

The way is now open for committees to introduce a rule that all discussion must be limited to the material in the Team-leader's report, of which the Trainer has had a copy for verification. This rule must be well-policed by individual committee members in the interests of fairness and accountability in criterion-referenced decision making.

4.5.7 How can we improve the reliability of Team-leaders' reports?

A concern that has been aired recently in our Deanery is that Team-leaders tend to couch their criticisms tactfully to the point of obfuscation (H Crawley, personal communication). If this is the case, then the more experienced committee members may be expected to be better at "reading between the lines" of the Team-leaders' reports.

A more homogenous format for Team-leaders' reports may help to ensure that committee members, especially new ones and those who attend infrequently, do not have to 'read between the lines' to interpret the quality of the training environment which the visitors found. If Team-leaders made it clearer when a practice does or does not meet the training criteria, then it is to be hoped that the reliability of these reports as an assessment instrument would be improved.

4.5.8 Conclusions from researching Team-leader reports and the assessment process

Team-leaders' reports are rated by different Trainer Selection Committee members with a high level of agreement in respect of the overall training quality. There is

Evaluating Process and Outcome in the Education of General Practitioners

higher concordance between the 'regular' adjudicators, than between 'occasional' adjudicators, and highest levels of agreement are seen for 'best-performers'.

Committee decisions on training practice accreditation correlate reasonably well with the quality of training as judged by individual team members viewing Team-leaders' reports.

There are examples of practices faring badly in terms of accreditation, when the mean rating of training quality by committee members, (considering the Team-leader's report before the committee meeting), had been high. Sometimes there was a wider range of ratings for these practices. Prior knowledge of the training, and experience at detecting important detail in the reports are possible factors, and there are some clues that seniority may sometimes have a bearing on the weight accorded to individual opinions in committee discussion.

The new framework of educational behaviours, which I have developed, and discussed in the previous chapter, should prove helpful in criterion-referencing the section of the Team-leaders' reports which address the quality of education.

This leads me to make some recommendations for change in practice, which are picked up in Chapter 6.

4.6 Reflections Phase IIb and IIc

The incorporation of new methods for assessing training behaviours into routine re-accreditation visits was an exercise in pragmatism and compromise. Balancing the requirements of the researchers for high-quality detailed data, against the needs of assessors for simplicity on an already over-crowded visit schedule, proved to be a tightrope act. How do the compromises look in retrospect?

The early decision to base novel assessments on existing methods in current practice, adapting to our purpose assessment tools that were already in use, was sound. In negotiations with Team-leaders and GP Trainers I encountered substantial resistance to any innovations, which could extend the visit or make it more complex.

4.6.1 Standard Scenarios reconsidered

I was disappointed not to be able to convince Team-leaders to incorporate standard scenarios, and in due course this may need to be revisited. Changes in the assessment process, and increasing confidence in the process of assessing educational behaviours will reduce the resistance to new methods *per se*. What remains valid is the point raised by Team-leaders that vignettes and standard scenarios are forms of assessment directed towards “preferred behaviours” and that GP Trainers (like any other subjects of assessment) will become conditioned to ‘playing the game’ in the desired way: the behaviours will be rehearsed and produced. My reflections on this are two-fold. Firstly that there is some evidence that behaviours evinced in response to vignettes closely resemble those seen in the ‘gold-standard’ assessment of behaviours by covert actors (Peabody, Luck *et al.*, 2000). Secondly, and possibly related, rehearsing desirable behaviours in the interests of performing well in assessment can, I believe, help to ingrain these behaviours into everyday practice (Holmboe and Hawkins, 1998)(something I observe as GP Registrars repeatedly review videotapes of their consultations and look for evidence of the behaviours prescribed by assessors).

The principal adaptation made to procedures on visits as a result of my research has been to train visitors to record evidence (in the form of ‘snippet’ jottings) during the course of interviews, video discussions, and inspections of records. The evidence is then reviewed in the course of the visit and a judgement made as to whether the preponderance of evidence supports one type of training behaviour or another.

4.6.2 Context specificity issues

There is little published work about the consistency of Trainers in their teaching behaviours. Work from undergraduate medical teaching would suggest that medical tutors stay reasonably true-to-form, and that it is only necessary to observe the teaching process on one or two occasions (Dolmans, Wolfhagen *et al.*, 1996). The very limited evidence from my research supports the notion of consistent behaviours – thus the Trainer who uses Socratic questioning to induce reflection in tutorials is likely to do so in debriefs and joint consulting, according to Registrar reports. Skills in feedback appeared also to be quite consistent. I still have concerns about context-

Evaluating Process and Outcome in the Education of General Practitioners

specificity in more complex behaviours such as completion of learning cycles, or contextualising family medicine; it remains possible that Trainers exhibit one behaviour when 'cued in' (for example in audit projects) but completely different behaviours in other contexts of everyday practice.

The process, of recording evidence and weighing it up, resonates with existing General Practice Training procedures. Each GP Trainer has to complete a Trainer's report as part of the Summative Assessment process, which each GP Registrar is obliged to undergo (Johnson and Hasler, 1997). Quite intentionally, we make reference to this common experience of Trainers and their peer assessors.

Our early experience during the pilot project had indicated that the tools were useful for recording the sort of evidence that we were after. Having learnt the dangers of 'leading questions' in interviews (Cohen and Mannion, 1994), and determined that evidence should be deduced from very general questions whereby the respondent had less cues as to the evidence we were seeking. This depended on an ability to record relevant information as it 'dropped out' in the course of a general interview, and for this to work well, two implications became apparent. Firstly, it is helpful to have another recorder present at each interview, rather than the questioner attempting to maintain spontaneous flow of conversation at the same time as noting evidence that emerges. Secondly, and very importantly, all assessors need prior familiarity with categories if they are to be able to notice relevant evidence during the course of interviews, and to record it against the appropriate category.

4.6.3 Conflicting interests of research and quality improvement

Therefore, training of assessors is critical to success. This in itself is not unproblematic, as wide dissemination of our research categories and dimensions amongst the peer assessors inevitably risked contamination of the research. Peer assessors are just that – Trainers who will themselves be subject to assessment. Thus, a Trainer, who has been helped to become familiar with the categories and dimensions of training behaviours, will no longer be a naïve subject when it comes to his or her turn for re-accreditation.

Evaluating Process and Outcome in the Education of General Practitioners

This is inevitable, and the decision to publish our research findings from phase one of the project (Peile, Easton *et al.*, 2000) was taken knowingly. The loss to research purity is more than balanced by the gain to dissemination of information, which we hope will lead to quality improvements in training behaviours - the ultimate objective of the whole research project. Change in behaviour is not achieved easily (Koutsavlis, 2001) and needs a joint focus on internal processes and external influences (Grol, 1997), of which assessment is one.

Experience in Phase IIb had shown that we were able to collect relevant information on all eight categories of training behaviours. At the time of rolling out the assessments, it remained to be seen if all these categories would prove useful in formative or summative assessment of GP Trainers. At the very least, I hoped to accumulate information on how Trainers perform in each category. This information was seen as important for individual feedback, also collectively for shedding light on where to place the emphasis in continuing education of the Trainers. In practice, Team-leaders commented that they found it helpful, when it came to feeding back to Trainers, to have evidence of their performance across the wide variety of categories. As in the pilot phase, no Trainer performed uniformly 'well' across all categories, so Team-leaders were able to offer evidence of relative strengths and weakness to each Trainer and suggest areas for possible improvement, as well as selecting areas for commendation.

4.6.4 No proof that preferred behaviours are better

The use of quality words in the preceding paragraph illustrates another problematic area of my work. To describe behaviour as good or bad, better or worse, improvable or commendable, is to suggest that we have established that the descriptors of 'preferred behaviour' and 'less helpful behaviour' have some established validity. As yet they do not. They rest on perceptions, which I hoped to validate in the final phase of this project.

However, much as in the paradigm of action research, I felt that I should not wait for 'proof' of benefit before working with research findings to improve practice. This is contentious, but pragmatic. If what we offer to Trainers as evidence of their training

Evaluating Process and Outcome in the Education of General Practitioners

behaviours has face validity to them as practitioners, and if this evidence suggests change, which seems to them productive and beneficial, then they are likely to change in that direction.

Much of the debate about OFSTED visits to schools centred on a similar dilemma, where formative feedback was based on somewhat subjective evidence in the context of a summative procedure (Fidler *et al.*, 1998).

Trainers welcomed the feedback on their teaching behaviours, but it should be remembered that this feedback was delivered by Team-leaders trained and skilled in the art of constructive feedback. Skilfully delivered peer feedback generally improves performance, (Costa and Kallick, 1993; Munson, 1998), especially if feedback is reinforced with mentoring (Freiberg *et al.*, 1987) in contrast to student feedback, which does not always improve teaching performance and may indeed lead to deterioration (Litzelman *et al.*, 1998).

4.6.5 Assessing longitudinal process in cross-section

A technical problem was the difficulty of making a cross-sectional assessment of a longitudinal process. GP Trainers work with Registrars over a whole year, and this period often sees progression in craft learners from being 'Dependent Learners' towards becoming 'Self-directed Learners'. The work of Gerald Grow (1991) suggests that it is helpful if the teaching style matches the need of the learner, changing along an axis between 'Authority, expert' to 'Delegator'. This is especially important when considering Category 1: 'Training or Education', as the needs of the novice are different from the needs of the advanced practitioner in this respect. The former, in say the first month or two, needs more of the behaviour that we label as 'less helpful' in order to become proficient. Once they have learnt some of the tools of the trade, then Registrars' needs change towards the more Socratic behaviours we label as 'Preferred' for the bulk of their time in the training practice. Heron (1999) explores options on the continuum of facilitator modes, from hierarchical through co-operative to autonomous, which apply equally well in one-to one teaching as in small groups.

Evaluating Process and Outcome in the Education of General Practitioners

It follows that to assess fairly across this dimension we need to try to find out not only how the Trainer behaves with respect to the Registrar currently in practice, at whatever stage he/she is in the training year, but also if and how the behaviour adapts for Registrars at different stages.

During Phase IIb the Standard Scenarios had proved useful for assessing behaviours which might not have been demonstrated on the visit. Judicious use of a scenario depicting a learner at a different stage of training (to the one we interviewed and saw in the video tutorial) enabled us to see evidence of how the Trainer adapts to different circumstances. We felt that we saw evidence of what the Trainer probably *does*, rather than the evidence of *what he/she says he/she does*, than we get when we just ask the question, “How do you adapt your training style for learners at different stages?”

4.6.6. Ethical Considerations

Our work was not without ethical dilemmas. It was determined early on that Research Ethics approval was not strictly necessary for the pilot visits, as there were no implications for patients. Since that time, the regulations have been tightened and all research carried out on NHS premises now requires approval. Having subsequently published draft guidelines on the ethics of Medical Education Research (Peile and Slowther, 2001), I would now have preferred to have sought ethics approval, as I believe that there are implications for the practitioners of having their training behaviours scrutinised, and of recompensing them for their time spent in assisting the research.

Even more challenging is the situation pertaining in Phase IIc, whereby the research proceeded without the informed consent of the participants. This is because the Directorate, which has the responsibility of determining the conduct of re-accreditation visits, decided that the new assessments would be incorporated in the routine visit procedures for a trial period (see Appendix 4). This situation has many parallels in education, where faculty has the power to determine changes in curricula or assessment procedures, without reference to Ethics Committees. The grey area is where these changes are researched or evaluated.

Ultimately, much of the justification of this work depends on whether it can be shown that certain training behaviours improve the quality of performance of trained practitioners, and whether assessing these behaviours leads to positive change in Trainers' performance. I will return to this question in Chapter 6.

4.7 Summary

Phase Two saw the evolution of a process to evaluate the actual training behaviours used by GP Trainers. This process was initially piloted on volunteers, and was subsequently performed on routine assessment visits over a trial period. Pragmatic compromises had to be reached between the objectives of 'pure' research and the feasibility of procedures, which depend on the efforts of assessors already fully committed on their primary task of preparing a report for summative assessment.

The complexity of the recording process, limited the value of the research findings from Phase IIc, when assessments were being carried out in the course of very busy, demanding visits by teams whose priority was the service task of assessment rather than the research endeavour.

In order to realise the potential of this work, both for assessment of Trainers and for formative improvement in educational behaviours, there needs to be a further stage IId, (Revision), which is discussed in Chapter 6.

CHAPTER 5
PHASE THREE RESEARCH
Collecting and Analysing Perceptions in the Assessment of Quality Performance Outcomes

5.1 Conceptual: Design Considerations Phase Three (Methods)

There are two research questions for Phase Three:

- What is the most practical means of collecting reliable data about quality of performance of GPs, in order to have an **outcome measure** against which to determine effectiveness of educational process?
- What is the best technique for relating process to outcome?

The objective of Phase Three is to look for meaningful outcome measures in terms of assessed competence and perceived performance of trained doctors in practice.

5.1.1 Attributes desirable in an assessment process

The desired attributes of an assessment process, whether of medical practice or educational practice are well defined (Van der Vleuten, 1996; McKinley, Fraser *et al.*, 2001) and are reproduced in Appendix 7.

5.1.2 Measures of quality doctoring

The obvious problem facing me was to decide on outcome measures that truly represent quality in a GP. The literature on quality in general practice is extensive (Hurwitz and Vass (2002); Marshall, 2002). Appendix 8 lists and gives the web-link to the RCGP searchable database of 39 topics, where quality standards have been determined. It is outside the scope of this thesis to engage fully with all aspects of quality markers in practice, but I propose to examine some of the problems facing me at the design stage and subsequently.

One of the leading British experts on quality, Professor Martin Roland, put his finger on the quality dilemma when he gave the 1998 James MacKenzie Lecture, "Quality and efficiency: enemies or partners?". He quoted James McCormick:

Evaluating Process and Outcome in the Education of General Practitioners

There seems to be a deep seated fear that techniques based on statistical measures would distort truth and are by definition unsuited to our examination of care. But surely if caring is of value, and that value is real, then it must be demonstrable.

To demonstrate may be less onerous than to measure, but Marshall Marinker (1992) famously encompassed some of the difficulties of demonstrating that reality:

The task of the hospital specialist is to reduce uncertainty, to explore possibility and to marginalise error. The task of the General Practitioner is to accept uncertainty, to explore probability, and to marginalise danger. Excellence in generalism is characterised by a superficial grasp of a very wide field of endeavour; problem solving is horizontal rather than vertical; thinking is as much divergent as convergent; the skills are not so much technical as interactive; there is room not only for the rational and explicit but also for the intuitive and implicit.

If we accept Marinker's dictum as epitomising the qualitative subtleties we wish to ascertain in looking for the good doctor, then we can look to Pringle and colleagues' (2002) recent article entitled 'Measuring "goodness" in individuals and healthcare systems' to set the standards for data.

Data should be accurate, measures appropriate, context adjusted for, and interpretation responsible and cautious.

These same authors argue for a mix of healthcare process and outcome variables to provide evidence of clinical competence.

Evaluating Process and Outcome in the Education of General Practitioners

Irvine and Donaldson (1993) published a review of quality improvement work in healthcare, and emphasised the importance of obtaining professional, managerial and patient perspectives, and of gaining the commitment of health professionals themselves.

Roland (1999) has done a lot of work trying to deconstruct the complexities of quality, and he points out that aspects of GP care most highly valued by patients are:

- Technical competence
- Communication skills
- Inter-personal attributes

The attraction of the measurable or the demonstrable is considerable, and indicators such as prescribing behaviours (Campbell, Cantrill *et al.*, 2000), referral rates (Dixon, Holland *et al.*, 1998); (Davis, 2001) and even mortality data (Frankel, Sterne *et al.*, 2000) have all been looked at as candidates, but all found wanting. For example, Roland (1999) found about half of factors controlling admission rates are outside the control of GPs.

Best evidence studies on guideline implementation sometimes use patient outcome measures for professional behaviour, for example:

- Percentage of patients reached
- Compliance with protocol (Katon, Rutter *et al.*, 2000)
- Patient outcomes/satisfaction (Scholle, Weisman *et al.*, 2000)

5.1.2.1 Quality Marker Options for this research

After reviewing the literature on quality markers, it was apparent that there is neither a comprehensive and widely accepted marker of quality performance applicable to a British GP, nor are many markers accessible without considerable effort on the part of the research participants, and this could be problematic.

5.1.3 Research participation has to be feasible in line with the constraints on GP time

This programme of research, as it turned out, was to be conducted at one of the busiest and most-pressurised times in recent General Practice history. As it was always going to be unlikely that GPs would voluntarily give up more than a few minutes to participate in research and in designing the 'quality' assessment, I had to plan on using information which doctor's could readily supply with minimum effort.

At the time of planning the research, it was clear that the government was going to introduce compulsory revalidation for all doctors, and, it seemed likely that this would include a personal portfolio of evidence.

My research strategy depended, in part, on connecting with individual doctors' preparation for revalidation: voluntary research participation would be perceived as helpful in the preparation for compulsory revalidation, I had hoped.

Practitioners' time in research participation would be rewarded with reports which would prove valuable as part of a revalidation portfolio.

5.1.4 The workload in General Practice

In Table 5.1, I set out the programme for my research against the current scene in General Practice. It can be seen that when I was trying to recruit practices to help with my research, there was an unprecedented combination of low morale, high workload, reduced workforce, shifting political and managerial environment, and uncertainty as to the political agenda. Undermanned practices were releasing partners to attend frequent meetings with the new Primary Care Groups, and GP educators and Trainers were struggling to meet their commitment. It was small wonder that some practices placed blanket restrictions on research participation in 2001, just when I needed them most.

At the time of research design, I was trying to secure evidence of quality markers, which would require minimal work on behalf of participating practitioners. As it turned out, I was unable to recruit anything like the intended number of practices.

Evaluating Process and Outcome in the Education of General Practitioners

Table 5.1: The Research Programme in context of General Practice Scene

	Research stage	Workload crisis	Appraisal/Revalidation	Political Restructuring
1998/1999	Enrolled at OBU (Oxford Brookes) Oct 1998 Research Design commenced Phase I started end 1998	Workload “constantly increasing” GPC recruitment & retention difficulties spiral worst workforce crisis for > 30 years <19% of SHOs choosing General Practice as career. New medical schools announced.	Jan 1999 The Health Bill – Clinical Governance and revalidation agenda Bristol Children’s Hospital Enquiry commences and public attention focussed on medical error and negligence	Grouping into Primary Care Groups (PCG’s) National Clinical Governance Support Centre (NCGSC) starts along with NICE (National Institute for Clinical Effectiveness) and CHI (Commission for Health Improvement)
1999/2000	NHS R&D fellowship started Oct 1999 Phase II a started Oct 1999 – Oct 00 Phase II b starts	Workload rising with secondary to primary shift Retirements exceed recruitment as generation of Asian practitioners retire: situation predicted to get worse till 2007.	Shipman scandal forces rethink on revalidation issues ‘Revalidation for clinical General Practitioners’ & ‘Good Medical Practice for GPs’ published	PCG’s looking at trust status Clinical Governance Agenda formalised. Government incentives for PMS (Personal Medical Services) groups erodes the long-established partnership base of General Practice as many doctors opt to become salaried rather than partners.
2000/2001	Phase II c starts Oct 2000 Phase III ethics approval, and starts April 2001	Scottish survey 56% of GPs low/v low morale 71% say morale lower than 5 years ago many looking to retire early Crisis of morale ‘endemic’ as 40% vacancies unfilled after 1 year.	Appraisal Scheme to start April 2002, but delays on revalidation. Appraisal by peers will need Personal Development Plan (PDP) but no patient or colleague perceptions	Primary Care Trust status for all PCGs General Medical Council (GMC) reform Medical Indemnity Claims spiral and one insurer closes to business Underperformance to be investigated in each region.
2001/2002	Phase III finishes April 2002 Analysis and writing up April to November 2002	GP Ballot – absolute majority prepared to resign National Survey of GP Opinion: 66% low morale 82% GPs “excessive work-related stress” 48% GPs intend to retire pre 60y old 28% seriously contemplating career change	Appraisal training cascaded out by NCGSC Not enough trained appraisers in place. Revalidation to start in 2007, but no decision yet on 360 degree inclusion.	CHI becomes CHAI (Commission for Health Audit and Inspection) National Clinical Assessment Authority (NCAA) starts Government authorises negotiation on radical new contract for General Practice.

Source for general data: Annual Reports of British Medical Association

5.1.5 The Case for a Perception-based Research Instrument

Since other quality measures were either too narrow in their focus, too contentious, or too demanding for the purposes of research collaboration, I turned to the literature on patient-based survey measures (Wensing, Jung *et al.*, 1998). Time and again I found myself returning to the refrain of the perception based instruments "*I know a good doctor when I see one*" (Gerteis, Edgar-Levitan *et al.*, 1993).

Although it seemed crucial to place a lot of emphasis on the patient perspective, as 'consumers' are often the best critics of a service (Coulter and Elwyn, 2002), I was also convinced that the self-assessment of doctors and the views of their colleagues were of equal importance as their perspectives can give complementary information. I therefore decided that a multi-source perception-based tool should be the mainstay of this work (Violato, 2003; Ramsey, 1993). I was attracted to the idea of balancing self-assessment (Gordon, 1991) with patients' assessments of their doctors (Wolf, Putnam *et al.*, 1978; Anderson, Rakowski *et al.*, 1988; Anderson and Dedrick, 1990; Rigby and Metzger, 1992; Schlegelmilch and Carman, 1993).

5.1.6 Research Instrument 1: 360 degree assessment

Of the multi-source perception-based instruments available, I wanted one which had a horizontal 360 degree format (King, 2002; Connors and Munro, 2001), as opposed to many used in industry and hospital medicine which have a more vertical format for colleague input (seeking the opinions of those colleagues that report to the subject, and those to whom the subject reports in the hierarchy). I selected 'Insight 360' (Edgecumbe Consulting Ltd), which gives a validated assessment of patient views, partner and colleague perceptions against self-perceptions (Griffin, Sanders *et al.*, 2000). Previous work with Insight 360 had shown that saturation is achieved with 15 responders in a class for any given question (Griffin, Sanders *et al.*, 2000) confirming the work of others in similar analyses of multisource feedback (Ramsey, Wenrich *et al.*, 1993; Violato, Marini *et al.*, 1997; Hall, Violato *et al.*, 1999).

In using Insight 360, I had to adapt an instrument which was designed as a Management Consultancy tool, giving vital feedback to doctors on areas to prioritise for improvement, to a research instrument. The company gave me access to their computer programmer to make the necessary adaptations to the software. I retained

the original validated categories, but removed items that were looking at the service offered by the practice, rather than the individual doctor, and adapted the wording on some of the patient questions to make them correspond to the doctor self-perception items. The patient questionnaires are attached at Appendix 9.

5.1.7 The case for a portfolio of Competence

As portfolios are becoming more familiar tools in practice (Snadden and Thomas, 1998), and as the GMC had trailed its intention to ask for portfolios of competence in the revalidation process, (GMC and DoH, 2002), it seemed reasonable to look at a portfolio as evidence of quality. There are a number of externally assessed pieces of evidence which some doctors can produce as evidence of quality performance.

Table 5.2: Examples of evidence of assessed competence, which could be used for a portfolio either for revalidation purposes or for use as quality indicators in research.

Summative Assessment in Vocational Training – four compulsory elements: Multiple Choice Questionnaire knowledge assessment Audit projects Video assessment of consultations Structured trainer’s report. Examination Performance: Membership of Royal College of General Practitioners Other postgraduate qualifications and diplomas Any external peer-reviewed assessments such as: Quality Practice Award (QPA), Membership by Assessment (MBA) Fellowship by Assessment (FBA) Published work Approval as a trainer Receipt of performance-related payments

As mentioned earlier, I could not expect participant doctors to engage in time-consuming work collecting evidence for my research, but at the time when the research was designed, it seemed likely that GPs would be preparing portfolios for revalidation, and thus I hoped to persuade my research subjects that preparing a portfolio demonstrating competence, rather than being extra work, would be part of the normal preparation for revalidation. By introducing research exercises that would help them with their revalidation preparation, I had hoped to ensure GPs’ co-operation with my research.

5.1.8 Research Instrument 2: Portfolio of competence

I decided to ask practitioners to supply evidence that they could detail with a minimum of effort, and which offered me some comparative data on quality markers.

Table 5.3: Information requested for Portfolio of Competence

- | |
|---|
| <ul style="list-style-type: none">• Undergraduate and school attainment (baseline)• Summative assessment passes and referrals• MRCGP performance• Other postgraduate qualifications and diplomas• Any external peer-reviewed assessments such as QPA, MBA or FBA• Published work• Approval as a Trainer• Receipt of performance-related payments |
|---|

The plan for analysis of this was that two experienced GP Educationalists would be asked to review each respondent's Portfolio blindly and independently. They would be asked to grade the doctor's Pre-VTS record and Post-VTS record as demonstrating one of the following levels of demonstrated competence: Below average; Average; Above average, in respect of a population of doctors similarly trained.

Where the observers disagree, the Researcher was to record the discordance and negotiate an agreed grading.

This mini portfolio was very much a compromise between desirable and feasible, but was aimed to provide a triangulation of performance against recognised evidence of competence.

5.1.9 The case for looking at reflection

In the Phase One work, the two messages that came across most emphatically from Registrars were the importance of modelling behaviours by the Trainer, and the long-term importance for the Registrar of learning how to reflect constructively.

Reviewing the transcripts a further time, it became apparent that, without exception, the former learners mentioned reflection in some form: either commenting on ways in which their trainer had helped develop the habit, or emphasising the importance to

them in later practice. Because reflection seems so crucial to learners, it seemed important to focus on this as an outcome, if possible.

5.1.10 Research Instrument 3: Reflectivity questionnaire

All participating doctors were asked to assess their Trainer's educational behaviours by recall, so that we had a subjective view on Category 3 (Space for Reflection) and other relevant behaviours. They were then asked to complete a simple reflectivity questionnaire. (Appendix 15) I had searched in vain for a suitable instrument to measure reflectivity in doctors.

(King and Kitchener, 1994) reviewed the literature on two of the most popular instruments which measure critical thinking: the Cornell Critical Thinking Test and the Watson-Glaser Critical Thinking Appraisal. Neither instrument is suitable for my purpose. Not only do they predominately deal with well-structured problems (in practice doctors have to respond mostly to unstructured problems), but both questionnaires are also hugely time-consuming.

The most promising work in the field of producing a tool to quantify reflective thinking has been done by (Kember, Leung *et al.*, 2000). In their article they described the development and validation of a questionnaire to measure the level of reflective thinking in nursing students. Working with my colleague, Dr Regina Conradt, and in association with Dr Anita Berlin, of Imperial College London, we decided to develop a pilot questionnaire for doctors, based on the same framework as Kember used. Although unvalidated, we had hoped to get enough responses in the course of the present research to enable us to begin the process of validation.

5.1.11 The case for looking at consulting

Another approach to outcome measures is to single out for attention that activity which epitomises the GP's professional practice. The consultation is at the centre of the GP's work and in my own practice, occupies just over half the teaching time. It is reasonable to deduce that if consulting behaviour is learned rather than innate, then measures of quality consulting will speak to the education the doctor experienced.

5.1.12 Research Instrument 4: Patient Enablement Questionnaire

In consulting at least there is a well-validated instrument, fit for purpose. Over a period of several years, Howie's group have been attempting to study the quality of interpersonal care in General Practice from the 'outcome' rather than 'process' standpoint (Howie, Heaney *et al.*, 1999). They refer to 'Enablement', as they believe this captures the main aims of patient-centred care, being improved understanding of illness and feeling of ability to cope following consultations. A third of the variance between doctors' mean enablement scores (based on 100 unselected adult consultations) can be explained by longer consultations and greater personal continuity of care (as measured by patients saying they know their doctor well). Doctors who enable more of their patients and enable them better, are those who offer their patients more time and greater continuity (Howie, Heaney *et al.*, 1999). I decided not to measure these contextual variables as, arguably, they are more a property of the group practice than the individual practitioner, and thus can relate only indirectly to past training.

However, the central aspect of Howie's work in constructing a Consultation Quality Index is the enablement score, which is based on 6 questions (Howie, Heaney *et al.*, 2000). I hoped to persuade interested research participants to ask 50 consecutive patients to co-operate in this exercise, as I saw potential to triangulate my other outcome measures against such a well-validated and relevant measure of quality.

5.2 Conceptual: Original Design Considerations for Phase Three: Analysis - MLM in Education Research

Plewis (1997) presents convincing arguments for the use of statistical models in educational research. I shall adopt his convention of distinguishing between *response* variables and *explanatory* variables, and I base my understanding of the way modelling makes assumptions explicit in the transition between complex real world educational processes and the essence that is captured in a statistical model on Plewis' simplistic equation : $DATA = FIT + RESIDUAL$. I interpret this as meaning dividing the data we collect into two types- that which contributes to the explanatory variable and the residual, which may just be 'noise'.

Table 5.4: Techniques that can be used to look at any level of an educational process

1 <u>Simple regression with categorical explanatory variables eg analysis of variance table:</u>			
	<u>df</u>	<u>Sum of squares (SS)</u>	<u>Mean Square (MS=SS/df)</u>
<u>Regression</u>			
<u>Residual</u>			
<u>Total</u>			
<p>The table generates two useful statistics:</p> <ul style="list-style-type: none"> • The ratio of the regression sum of squares to the total sum of squares is R^2 • The ratio of the regression mean square to the residual mean square has an F distribution, (assuming the residuals are normal), with the corresponding degrees of freedom, enabling the p value to be calculated <p>These techniques are useful for comparing the data from very few high-level units, but models looking at, for example, 30 schools requiring 29 dummy variables would be cumbersome (Plewis, 1997).</p>			
2 <u>Multiple regression techniques</u>			
Plewis explores the use of such techniques in education			
<ul style="list-style-type: none"> • With continuous explanatory variables (especially useful in examining Longitudinal data) • With categorical explanatory variables a <i>main effects model</i>, models which include terms for interaction, (two-way or greater) • With mixed explanatory variables - the ANCOVA model (Plewis, 1997) 			
The order in which variables are examined in multiple regressions is important, and choices include:			
<ul style="list-style-type: none"> • Standard multiple regression • Sequential (hierarchical) multiple regression • Statistical (stepwise and setwise) regression 			
<p>The problem with conducting even quite sophisticated regression techniques at a single level is that within group regressions can differ from between group regressions and these differences could correspond to important educational processes.</p>			
3 <u>Canonical R</u>			
Canonical correlation is used to assess the relationship between a set of several continuous response variables and several continuous explanatory variables			
Again the same limitation applies in that the researcher has to choose at which level to analyse the variables, or risk falling into the trap of ecological fallacies by analysing aggregated data across different levels.			
4 <u>Multiway Frequency analysis</u>			
This technique can be used to assess relationships between discrete variables where none is considered a response variable. The technique becomes known as logit analysis when one variable is considered a response variable with the rest serving as explanatory variables. Logit analysis can usefully be built into multilevel models.			

5.2.1 Options for Statistical analysis

Consider the alternatives for making predictions based on the relationship between variables. As Tabachnick and Fidell (1996) explain, the choice of method depends on the number and nature of variables, and whether any of the explanatory variables are best conceptualised as covariates (Table 5.4).

There are a number of advantages to Multilevel Modelling (MLM), which are displayed in Table 5.5.

Table 5.5: Advantages of MLM

1. MLM can handle dummy variables for a significant number of categories (eg schools)
2. Within group regressions can be examined separately from between group regressions to look for differences that could correspond to different and complementary educational processes.
3. MLM is capable of examining cross-level hypotheses.
4. In MLM by taking account of **all** the variability in the data, the standard errors of the regression coefficients are correctly estimated, whereas the precision of estimated coefficients from single level regressions tends to be over-estimated (Aitkin, Anderson *et al.*, 1981) as reported by (Plewis, 1997).
5. MLM enables researchers to address questions which were previously unanswerable, - Plewis gives the example of whether the relation between attainment and curriculum coverage varies from teacher to teacher, and *why* this might happen (Plewis, 1997).
6. MLM avoids the pitfalls of data aggregation and ecological fallacy because researchers do not have to make a decision about 'at what level' to analyse - in MLM data should be analysed at all levels.

5.2.2 MLM working in parallel situations to my own research

Goldstein, Rasbash *et al.* (1993) looked at the contribution made by the education in the schools in question to the examination performance of the pupils. The authors' contention is that the technique of MLM enables them to isolate this contribution sufficiently to study the variation between schools. They state their conclusion that, "*...few schools can be separated reliably. In particular, no fine rank ordering of schools legitimately can be produced.*" They set out to address outstanding issues in developing "methods for comparing schools and other institutions on the basis of the achievement of their students."

Evaluating Process and Outcome in the Education of General Practitioners

After transforming the variables using normal scoring to conform as closely as possible to multivariate normality, they proceed to explore models to fit the data and they report the two separate models found to give the most satisfactory fit. In their first model, Goldstein's group are able to make several observations at the student level, "The effect of school gender is small and the differences are about the same order of magnitude as the estimated standard errors. There seems to be a small advantage for those attending Roman Catholic schools. Girls do better than boys."

An important contribution is made by constructing confidence intervals for the estimates of residuals. They chart the approximate 95% confidence intervals and find a very considerable overlap of intervals, which suggests that it is not possible statistically to discriminate easily between schools. In particular, there are no natural division points in the sequence of estimates to allow classification of schools into homogeneous subgroups.

In the fixed part of Goldstein's second model it is possible to look at the effects of gender on performance, (minimal). The lower achieving intake students had smaller variance in this model just as they had in the analysis of total score in the first model. The way in which Goldstein *et al* modelled the relevant variables at the level of the student and the school, so focussing on the contribution made by the schools to pupil performance, induced me to incorporate much of their methodology into my work, looking at the contribution of a training practice to the later performance of a doctor.

5.2.3 Multilevel models for school and teacher effectiveness

If the focus of the research is on *progress* rather than on *attainment*, then we need a model to incorporate a hierarchically structured two-level dataset, with a continuous response and a continuous explanatory variable as: $Y_{ij} = b_0 + b_1x_{ij} + u_j + e_{ij}$ where x_{ij} is an earlier measurement on pupil i in school j .

The random intercepts can be plotted, as attainment on the first occasion on the x -axis and attainment on the second occasion on the y -axis to show a different intercept for each school.

A highly important finding concerns the random effects in this model. As in the example given by (Plewis, 1997)) the variance terms for between school components and between pupil components fall dramatically. This is because “for pupils, the best predictor of present attainment is past attainment, and for schools, some of the differences in attainment at one point in time are likely to be accounted for by earlier attainment differences.” (This underlines the importance of *Intake Differences*.)

In the medical context, MLM techniques have the capacity to look at the contribution individual GP Trainers make to the development of individual Registrars.

5.2.4 Multilevel Theories

The design of a good statistical model depends on a deep understanding of the effects we wish to measure. “If there are effects of the social context on individuals, these effects must be mediated by intervening processes that depend on characteristics of the social context.” (Hox, 1995). He goes on to say,

In multilevel problems, decisions about group membership and operationalizations involve a wide range of theoretical assumptions, and an equally wide range of specification problems for the auxiliary theory. When the number of variables at different levels are large, there is an enormous number of possible cross-level interactions. Cross-level interaction effects between the individual and the context level require the specification of some process within individuals that causes those individuals to be differentially influenced by certain aspects of the context.

He points out that, “another theoretical area that has been largely neglected by multilevel researchers is the influence of individuals on the group.”

Once the model has been designed on theoretical lines, there comes a point where running the model repeatedly begins to provide pointers to where theory needs to be reconsidered. There is thus an iterative or spiral process between model and theory.

5.2.5 Classes of Multilevel Models

Two classes of multilevel models predominate in education: multilevel regression models and multilevel models for covariance structures. I shall concentrate on the former class, which is described by Goldstein (1995).

Evaluating Process and Outcome in the Education of General Practitioners

Multilevel regression models include models which have variously been described as the 'Random coefficient model', the 'Variance component model', and the 'Hierarchical linear model'. The assumption in a simple multilevel regression model is that there is a single response variable measured at the lowest level and explanatory variables at all existing levels. The concept is of a hierarchical system of regression equations.

5.2.6 Hierarchical nature of educational data in MLM

The most usual treatment is to use hierarchical levels starting at the level of the student and then 'nesting' pupils within classrooms 'nested' within schools e.g.: Level 3 for school; Level 2 for teacher/classroom; and Level 1 for the student.

Many other possibilities exist. For example to treat multiple variates within each student as level 1, as Goldstein *et al* (1993) do, in research on schools' examination results. Also longitudinal data can be handled by treating different times as different levels.

The variables can be defined at any level of the hierarchy (Table 5.6). They may be measured directly at their natural level, and by aggregation, moved to a higher level, (Hox gives the example of computing the school means of the pupils intelligence scores), or by disaggregation moved to a lower level (here the example Hox gives is of assigning to all pupils a variable that reflects the denomination of the school they belong to) (Hox, 1995)).

Hox makes the point that for the purpose of analysing multilevel models, it is usually not important to assign each variable in the scheme. The advantage is conceptual.

5.3 Choices to be made in my research

The objective with the performance scores on different items in the Insight 360 was to perform factor analysis so that I might get a data reduction to about five factors each with an Eigen value >1. By using an appropriate algorithm for factor extraction, the factor loading would enable us to give precise values to a new variable score.

Table 5.6: A scheme for levels in MLM (Hox, 1995)

There is a choice in the ‘black art of statistics’ to adopt one of two approaches to factor analysis across different perspectives. The first approach is to factor analyse each perspective individually and then to perform a qualitative analysis of the difference between perspectives in the factor analysis. This would illuminate differences in the diversity of the perspectives. An example of unexpected difference in my work is the gender difference that emerges in self-perception by doctors. The second approach is to factor analyse each item across the perspectives and then perform analysis of variance. This approach looks for similarities not differences, and may be more appropriate, for example, in looking at different patient perspectives.

Insight 360 invites respondents to grade items on 2 separate Likert scales - one for ‘Importance’ of the item and the other for the doctor’s ‘Performance’ on that item. There are 2 potential advantages to including the ‘Importance’ data. Firstly, respondents feel more involved given an opportunity to say what is important to them. Secondly, the importance data may give some information about the culture in which the responses are being given. I therefore intended to include the importance data but

analyse it separately to look at cultural information following the multi-variate analysis on the performance data.

5.3.1 Is MLM the best approach to adopt?

Studying in an Institute of Education, I first encountered MLM in the analysis of school performance, and immediately considered that the technique might lend itself to my purpose. I was very struck by Plewis' (1997) observation that,

As soon as we allow for the possibility that slopes vary randomly, we are also introducing the possibility that there is no longer a single rank order of schools in terms of their effectiveness. Instead, some schools could be more effective for low attaining pupils, whereas others could be more effective for high attainers.

I had seen this powerfully demonstrated by Goldstein's paper, and I realised that it is also highly important for my research to look for complexity in the ways in which Trainers affect the subsequent performance of their GP Registrars.

Not being a statistician, it was possible that I was allowing my enthusiasm for a technique to drive the methodology, rather than allowing methodology to 'follow where metaphysic leads' (Harre, 1970). I therefore went on a course in York to learn about MLM applications in medicine, and had further debates with medical and educational statisticians, which convinced me that MLM was appropriate for my research – indeed only an analysis that can ascribe variance to different levels can possibly help to untangle the causality of complex educational processes.

5.4 Designing the multilevel model

The considerations around a suitable model for my purpose were how many levels to use, and what to assign at each level, what constitutes response variables and explanatory variables, and constant variance terms.

5.4.1 Considerations regarding levels

My original design intention, assuming a large number of trained Registrars, many from the same practice and many from the same VTS take part, was to perform a 3-level analysis: Level 1 for the trained doctor; Level 2 for the training practice; and Level 3 for the District VTS scheme.

I considered the possibility of making the District Day Release Scheme a level two variable rather than a separate level with different training practices nested within it. I prefer the three level model as it seemed more correct conceptually to isolate on a different level the contribution made by education to the doctor's later performance from that made by the training practice.

Depending on my sample, it would be relatively easy to incorporate additional levels into the model. For example, many practices have had different Trainers operating simultaneously or consecutively, and a useful device might be to make Trainers a separate level nested within training practices. This illustrates the iterative process of conceiving a model and then reconceptualising to accommodate data.

5.4.2 Designing the multilevel model: Response Variable

Performance scores (factor loaded), from Insight 360 are prioritised, but I also intended to replace this by the post -VTS record of demonstrated competence, (graded into 3 discrete categories) in a second, (otherwise identical) model. This would allow me to examine the null hypothesis: "that educational characteristics of training practices predict better for competence, (as demonstrated by recognised achievements in examinations and external assessment), than for performance, (as demonstrated by Insight 360 factor loaded scores)." Only if able to predict day-to-day performance would my work have achieved its full objectives.

5.4.3 Designing the multilevel model: Explanatory Variables

I hoped to have enough data to explore five explanatory variables (Table 5.7). It would be expected that discrete variables would be the more easy to relate.

Table 5.7: Planned explanatory variables in my research

<ul style="list-style-type: none">• Training practice (old Method) mean global score of committee rating Team-leader's report (<i>Continuous</i>)• Training Practice (new method) Scores on selected 'Educational Factors' (<i>it was envisaged that this would probably be Continuous data.</i>)• Pre-VTS record (intake attainment) graded into 3 <i>discrete</i> categories• Gender (<i>discrete</i>)• Time since completed training (<i>Continuous</i>)

5.4.4 Designing the multilevel model: Complex variation at level one

An issue left to consider in the light of the data was that of whether to assume a constant variance term σ^2_e at level one. By making the coefficient of a level one variable random at level two, the variation at level two varies according to the level one variable. By making the coefficient of a level one variable random at level one, the level one variance is no longer constant, and heteroscedascity can be built into the model (Goldstein, 1995; Plewis, 1997). An example of this is 'Time since training'. If a scatter plot suggested that there might be an effect of level one variance by 'Time since training', then I could test this by making 'Time since training' random at level one and including two further random parameters at level one (a variance and a covariance), to see if the fit of the model improves.

5.5 Measurement in Phase Three

In this section I shall look at the way the Insight 360 questionnaires performed and also at the use of MLM in the analysis. The other intended outcome measures will be mentioned, but severe recruitment problems meant that there is little data to be analysed for these.

5.5.1 Recruitment problems

At the time of the design stage, (1999), the indications were that revalidation would be introduced by 2002 and that a portfolio of evidence, including probably a multi-source feedback, would be a required component. Thus, I anticipated that doctors would welcome the chance to take part in the Insight 360 degree exercise, as this would give them, free of charge, a valuable component of their revalidation portfolio. As it turned out, we were to be confounded by the delays in the system. The programme for revalidation was not announced until 2002 (GMC and DoH, 2002), and the first revalidations do not take place for years to come. It proved exceptionally difficult to persuade doctors to take part in this sort of research, at a time when GPs in the UK were under more time pressures than ever before (GPC, 2001).

Less than 1 in 8 of the doctors we approached (all doctors trained in Oxford Region over the past 10 years) responded to 2 invitations to take part. The most popular option was for whole partnerships to take part together and so I promoted this

vigorously. Ultimately we were able to persuade only 33 doctors from 7 partnerships to take part. Our data comprised 601 perceptions across all 4 categories. The original target of recruiting three hundred doctors could not be based on appropriate power calculations as these do not apply in MLM. Only trial runs of a model will confirm if there is adequate data.

5.5.2 Insight 360 Perceptions

The Insight 360 patient questionnaire is attached as Appendix 9. The self-perception questionnaire and colleague questionnaire are very similar and cover 6 more items. The respondents were asked to give their perceptions on the target doctor in the 4 categories retained from the validated management consultancy tool (Table 5.8).

Table 5.8: Insight 360 categories

Category 1 = ‘Patient Care’	Category 2 = ‘Happy’
Category 3 = ‘Professional’	Category 4 = ‘Up-to-date’

5.5.3 Responses to questionnaires

Nearly all target doctors rated themselves across all 40 questions. Two doctors did not answer questions: 14, 33 or 38. Patients as well as colleagues varied more in the acceptance or refusal of a question. It is worth looking at the non-responses as there is food for thought regarding questionnaire development. Overall, the unanswered questions in the different categories are summarised below.

Table 5.9: Non-response items

Question		Patients n=255	% Patients, who did not answer all questions in category	Colleagues n=353	% Colleagues who did not answer all questions in category
		Number of questions not answered in each category		Number of questions not answered in each category	
cat 1	Patient Care	196	6.41	380	8.97
cat 2	Happy	199	11.15	218	8.82
cat 3	Professional	546	21.41	1083	19.17
cat 4	Up-to-Date	398	31.22	393	22.27

Evaluating Process and Outcome in the Education of General Practitioners

In Appendix 10 the individual question responses are tabulated, using the same colours as in this summary table.

In the ‘Professional’ category, patients were only asked 10 questions. Six of the 16 questions in this category that were asked of self and colleagues, were deemed inappropriate for patients, and these questions were not taken into account in calculating ‘Professional’ non-responses.

Most patients were comfortable in answering questions about how the doctors interacted with patients, but felt less comfortable in judging what a doctor intended to do. A lot of patients did not feel happy answering questions about doctor-staff cooperation or potentially threatening questions. Patients felt more comfortable in answering questions in category 1 than in the other categories. However, even in category 1, two questions were not answered by more than 25% of the patients, presumably because of lack of personal experience of the situations to which they related (emergency care and telephone consultations). Only a little over half of all patient responders felt they could answer a question about “the doctor responds to complaints effectively”, again reflecting the fact that complaints in General Practice are still infrequent.

Colleagues seemed most comfortable answering more general questions related to the patient-doctor relationship. Slightly fewer colleagues answered questions about the quality of doctor’s work, or about interaction with other colleagues. Colleagues often declined to answer questions about doctor’s involvement in audits or in other personal development of the doctor, and this may have been because their perspective on the doctor’s work did not cover these aspects.

5.5.4 Examining the data

The data reliability was high (see Table 5.8), indicating a high degree of consistency for the measurements, but unfortunately this also reflects the compactness of our data, which also compounded the difficulty of examining variance.

Table 5.10 Data reliability

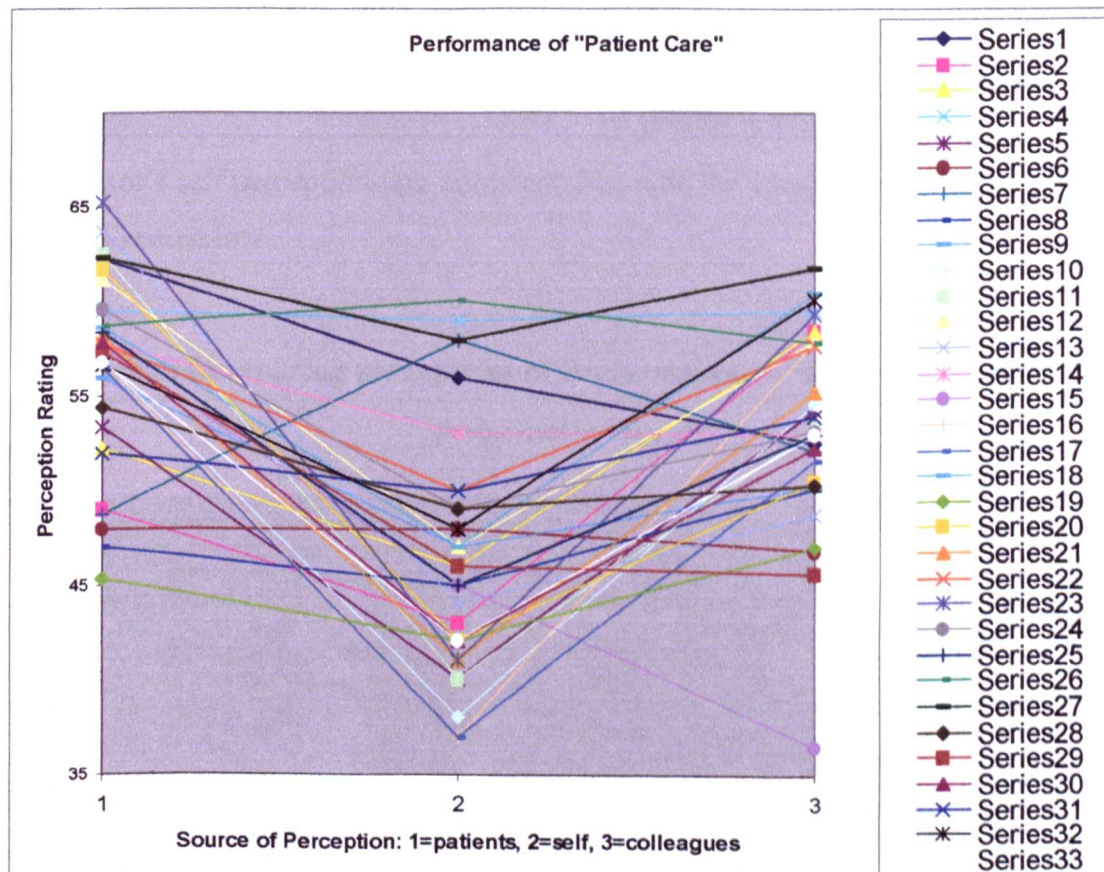
The Cronbach Alpha reliability co-efficients for our data were:		
Patient perceptions	alpha = 0.7401	n of cases = 331
Self perceptions	alpha = 0.8401	n of cases = 33
Colleague perceptions	alpha = 0.8474	n of cases = 237

5.5.5 How do perceptions vary?

The four aggregated categories ('Patient Care', 'Happy', 'Professional', and 'Up-to-date') are discussed separately. For each doctor we take the mean of the individual rating group, colleagues and patients, respectively. Self Perceptions are shown in the middle column of each of the 4 following tables; with patients' mean ratings to the left and colleagues' mean ratings to the right. Some interesting observations emerge.

As far as 'Patient Care' is concerned, the doctors tended to be harder on themselves in their ratings, than their patients or colleagues rated them, and the slope diagram suggests that, in the main, the latter 2 groups were similar.

Figure 4: Comparing perceptions of Performance in "Patient Care" category
(each 'Series' represents an individual doctor.)



A similar pattern emerges for the category 'Happy'. The absolute values are tightly grouped, but again, patients and colleagues are similar in the way they rate the doctors. This data compactness prompts one to look at the outliers to see if there is anything revealing about them. Here, the obvious outlier is Doctor 15, in respect of

colleague perception, so let us look at the raw data scores in respect of this doctor and also look at the slope diagrams to see if this doctor stands out.

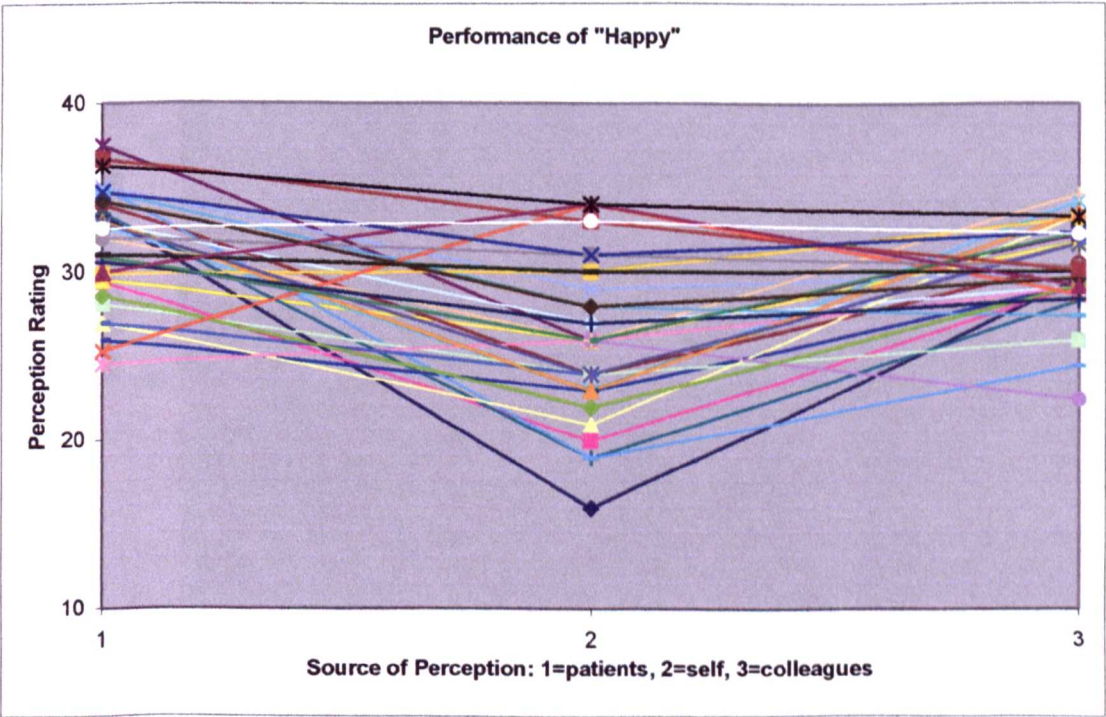
For ‘Patient Care’, Doctor 15 scores close to the mean of all target doctors in terms of both patient perception and self-perception but, as noted, colleagues rate this doctor well below the mean levels, as they do for all 3 other categories of performance, especially ‘professional’. Patients, on the other hand see the doctor as close to average in all respects except ‘Up-to-date’.

Table 5.11: Perception scores for a possible outlier: doctor 15

	Patients	Self-Perception	Colleagues
PATIENT CARE	57.6 (mean = 56.7)	45 (mean = 46.5)	36.5 (mean = 53.6)
HAPPY	30.8 (mean = 31.6)	26 (mean = 26.3)	22.5 (mean = 30.2)
PROFESSIONAL	35.7 (mean = 38.9)	44 (mean = 50.1)	35.0 (mean = 59.5)
UP-TO-DATE	14.9 (mean = 17.4)	16 (mean = 17.7)	15.0 (mean = 19.6)

The doctor’s self perceptions are unremarkable, with the exception of ‘Professional’, which is scored low.

Figure 5: Comparing perceptions of Performance in “Happy” category



Evaluating Process and Outcome in the Education of General Practitioners

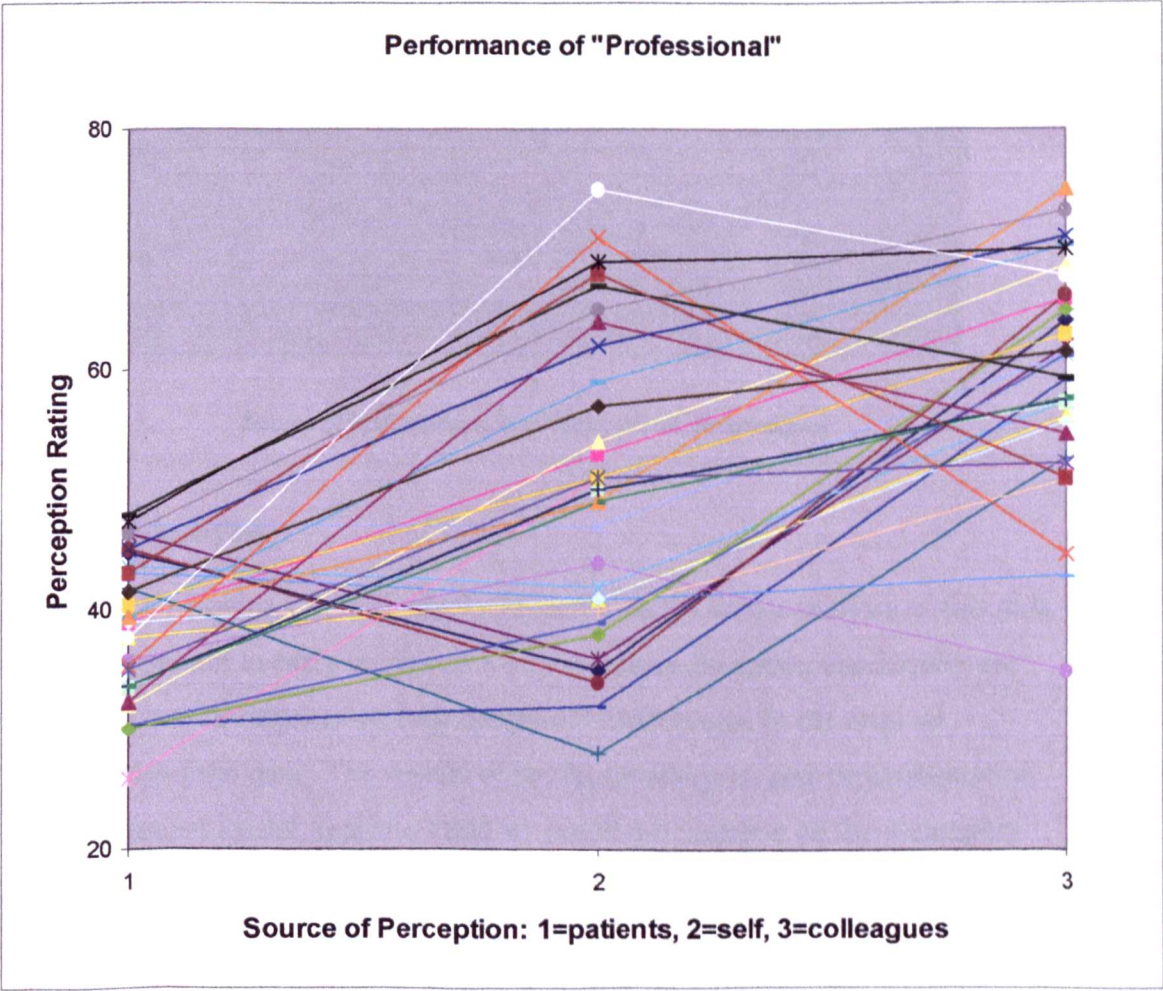
Examination of the slopes in ‘Happy’ shows that again doctors tend to rate themselves harder than their patients or colleagues rate them, and we see again that doctor 15 has the lowest rating by colleagues. But there is another doctor who interests me here, and that is doctor 22, who, unusually rates self significantly better than do patients or colleagues.

Table 5.12: Perception scores for a possible outlier: doctor 22

	Patients	Self-Perception	Colleagues
PATIENT CARE	57.8 (mean = 56.7)	50 (mean = 46.5)	57.7 (mean = 53.6)
HAPPY	25.3 (mean = 31.6)	34 (mean = 26.3)	28.7 (mean = 30.2)
PROFESSIONAL	35.3 (mean = 38.9)	71 (mean = 50.1)	44.7 (mean = 59.5)
UP-TO-DATE	30.0 (mean = 17.4)	21 (mean = 17.7)	16.7 (mean = 19.6)

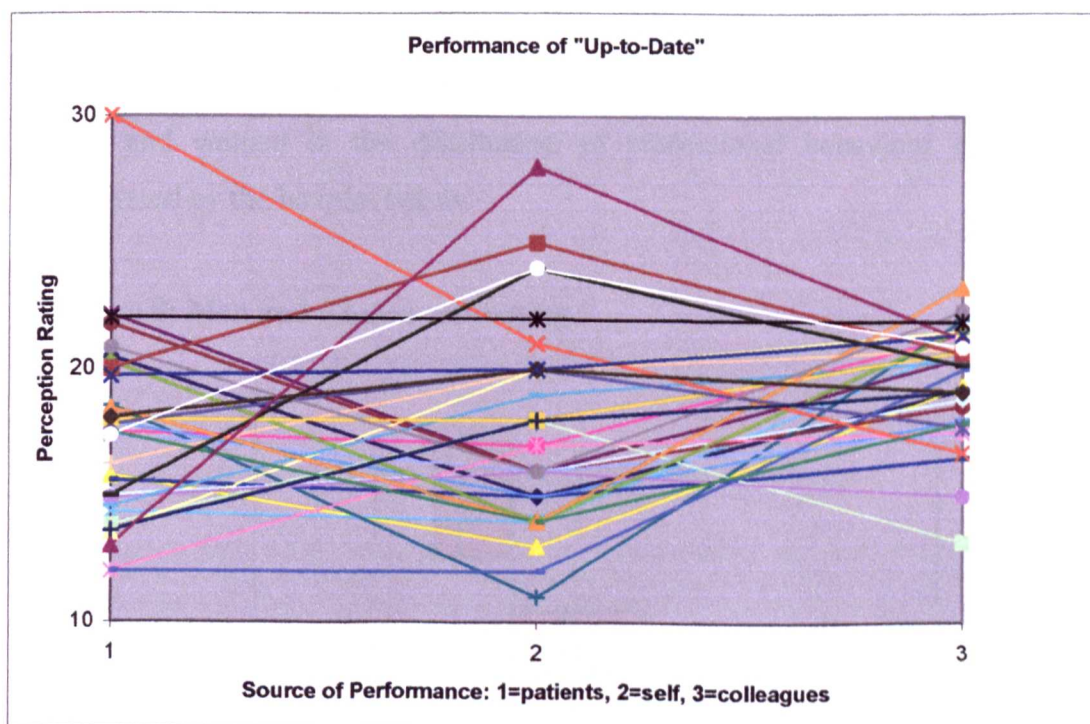
The impression of a higher than average self-rating on ‘Happy’ is confirmed, but look at how the same effect is seen (to an even greater extent) in ‘Professional’, where colleagues rate lower than average.

Figure 6: Comparing perceptions of Performance in “Professional” category



The slope diagram for 'Up-to-date' shows doctor 22 as an outlier, as previously mentioned. The general pattern, however, resembles 'Professional' in its rhomboid shape, with the doctors being more widely separated in their self-perception than the patients. There were a large number of neutral and non-responses on the 'Up-to-date' dimension, confirming the impression that patients found this dimension very difficult to assess. But almost an equal number of doctors rated themselves better or worse than their colleagues rated them in this respect.

Figure 7: Comparing perceptions of Performance in "Up-to-date" category



5.5.6 Factor Analysis

My colleague, Dr Regina Conradt, kindly helped with the factor analysis of this data. I felt it was important to carry out factor analysis before accepting uncritically the categories into which Edgecumbe had assigned the questions, in the original consultancy use of the data. The results of the factor analysis, and experimentation with a 3-component model, indicated that we could not improve on the 4-category model, which was employed in this work.

5.5.7 Looking at the data with MLM

I am very grateful to Dr Vanessa Simonite for her help with refining a multilevel model to interpret the data, see Appendix 11. Appendix 12 contains information on additional runs of multilevel models.

The value of MLM is that it enables us to explain where variation arises and the first question was how do we explain variation between doctors in their self-perception. It was immediately apparent that gender is one explanation of variance:

The estimates show that, compared to female doctors, male doctors' self assessment scores are significantly higher on average ($\chi^2=15.334$, $df=1$, $P < 0.0001$) and have significantly higher variation ($\chi^2=5.396$, $df=1$, $P = 0.02$). The tests are based on the change in the log likelihood when each term is excluded from the model, other things being equal. The difference between men and women in the distribution of professional behaviour ratings is illustrated by the boxplot below:

Figure 8: Men and women self-rating



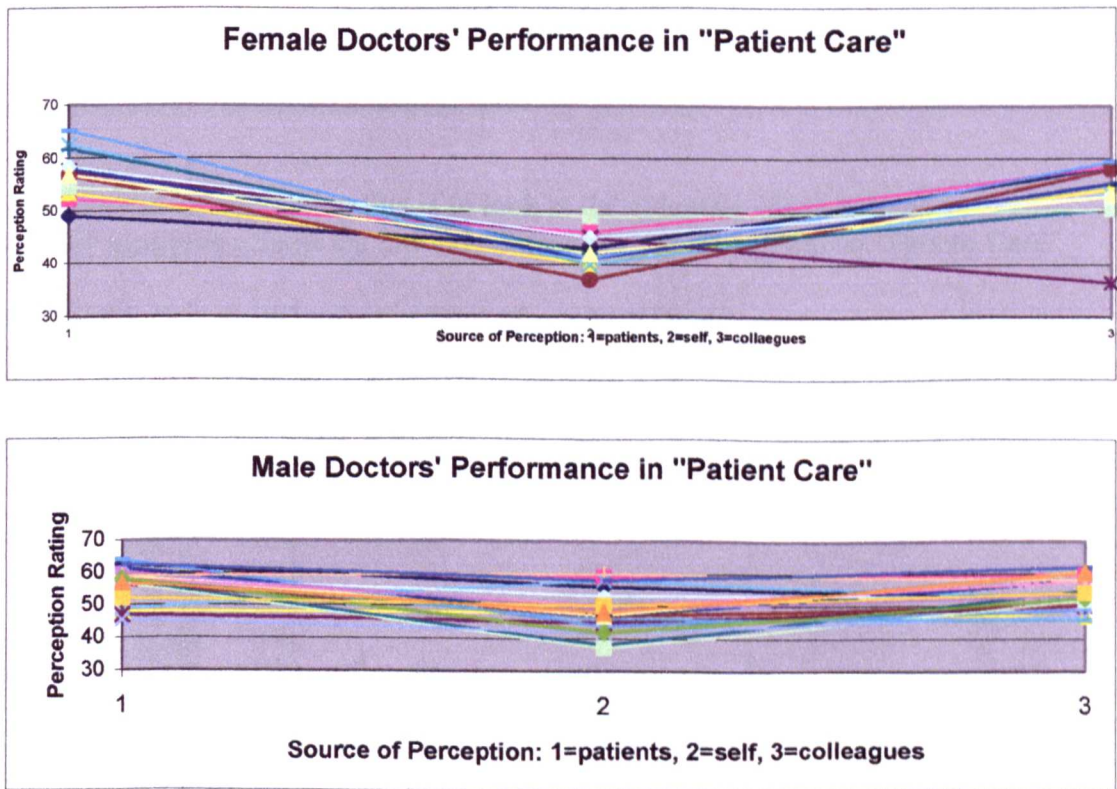
Finding such a difference poses a difficulty. If male doctors tend to see themselves differently from the way that female doctors see themselves, there are two possible explanations. The first is that both are perceiving themselves with similar accuracy; thus the differences reflect differences in actual performance. The alternative, and in my view, more likely explanation is that it is in the perceptions and not in the performance that the sexes differ.

Evaluating Process and Outcome in the Education of General Practitioners

Relating the multi-level model to my conceptual model (Appendix 2.4), this is an example of a *systematic* effect on the personal shadow of perception, if females are harder on themselves than males. It is important to try to understand which explanation is more likely to be correct as this affects how we compare self-perceptions to the perceptions of others. If males and females view themselves comparably, then their performance can legitimately be analysed together. If, on the other hand, the self-perceptions are filtered through different ‘gender lens’, then it may be more appropriate to make comparisons of self perception against patient and colleague perception in gender-separated groups.

The next two figures show how the slope diagrams look, separated by gender, for the dimension ‘Patient Care’. With the exception of the single outlying female doctor rated poorly by colleagues, the visual impression is that patients and colleagues rate the doctors similarly, irrespective of gender. This suggests that there is a tendency for females to self-rate lower, although others see them performing at the same level as their male counterparts. This impression is given credence by further modelling (*vide infra*).

Figure 9: Male and female performance and perceptions compared

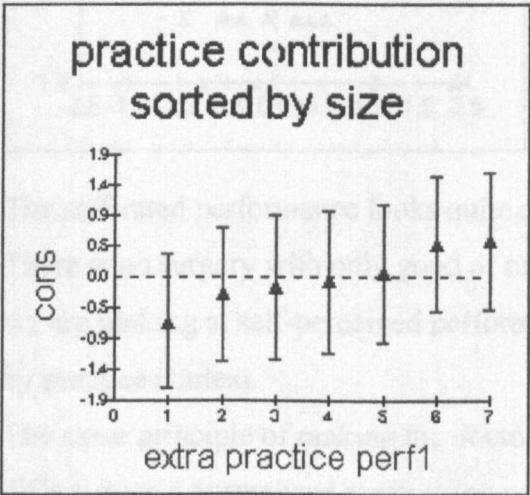


With my colleague, Dr Regina Conradt, who undertook the bulk of the modelling work, I resolved to assess the utility of MLM as a tool in developing this research methodology from the pilot stage to large-scale use. The formulaic modelling is included in appendix 12.

5.5.7.1 Doctors’ self-perception in ‘Patient Care’

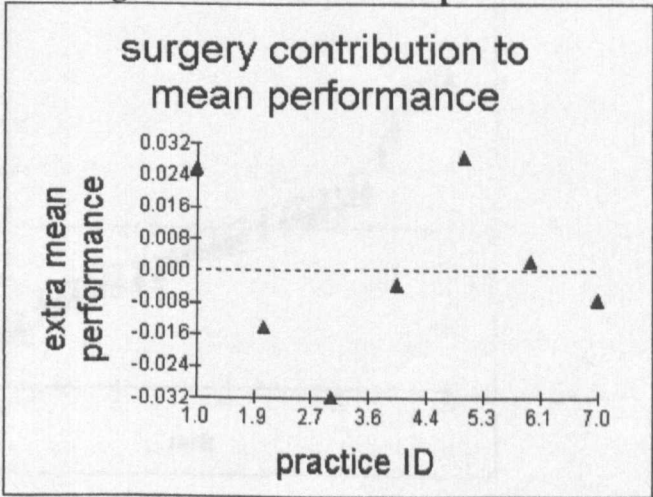
The self-perception of the doctors appears to be independent of the practices to which they belong. There is enormous overlap in the confidence intervals, in Figure 10 suggesting that, for ‘Patient Care’, there is no difference between surgeries.

Figure 10: Ranked contribution to variance by different practices



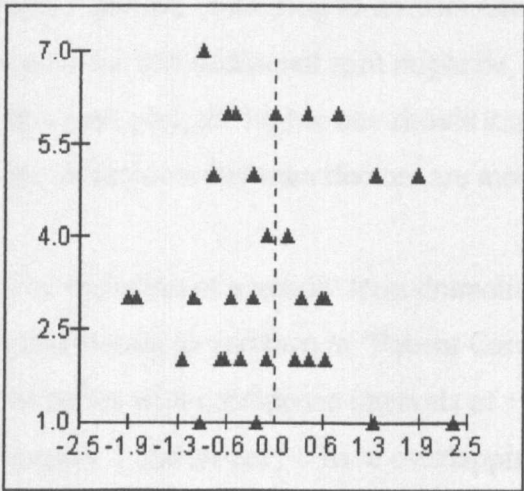
Another way of expressing this is to look at the extent to which belonging to any surgery grouping affects the individual doctor’s self perception on ‘Patient Care’.

Figure 11: Surgeries’ enhancement of performance



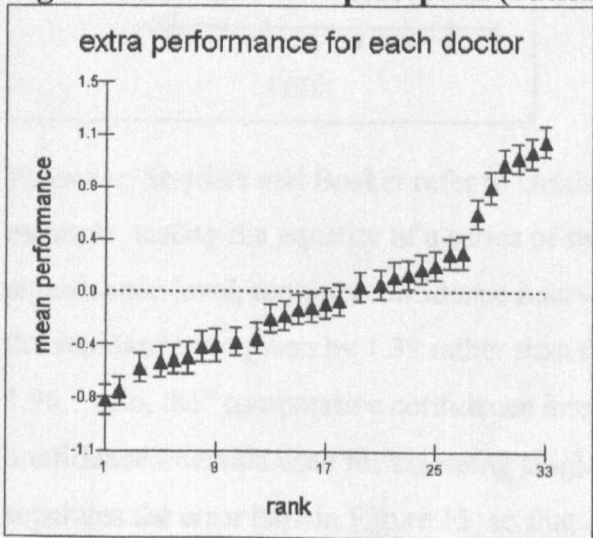
A single level model is adequate until covariates are added, such as gender or Registrar, when a level one model may no longer be sufficient to explain the data. Figure 12 shows the 33 doctors' self-ratings within different surgeries (y) plotted against normalized data of 'Patient Care' (x).

Figure 12: Doctors' self perception (Patient Care) grouped by surgeries



The self-rated performance looks quite evenly distributed between the surgeries. There is no surgery with only good or only poor doctors, confirming the finding that we are looking at self-perceived performance independent of significant confounding by practice context. The same principle of ranking the doctors by the extent to which their self-rating differs from a normalized mean using mean can be used to look for significant differences between doctors.

Figure 13: Doctors' self perception (Patient Care)

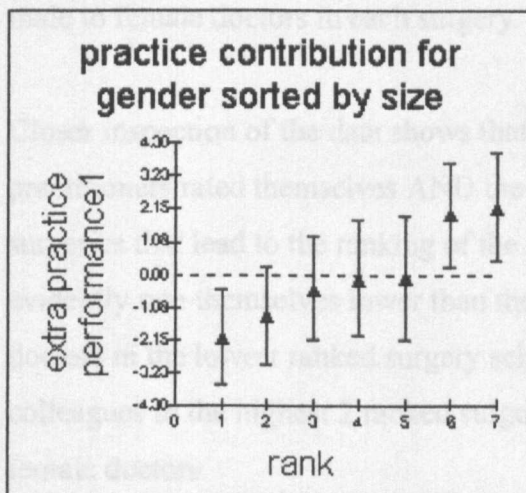


The plot seems to indicate that there are 4 sub-populations of doctors.

6 doctors (rank 33-28) considered their performance to be outstanding compared to other doctors' self-perceptions, and indeed there is no apparent overlap with the remainder of the population. At the other extreme there are two doctors who value themselves less highly than the majority of the population. The remainder can be split into 2 groups, consisting of doctors ranking 3-11 and 12-27, respectively. What the cause for this additional split might be, is unclear. Note that the higher the slope of this rank plot, the higher one should expect the categorisational power, in other words the differences between doctors are more marked.

The inclusion of a gender term dramatically affects the rank plot for practice contribution to variance in 'Patient Care'. Figure 14 shows the residuals of the 7 surgeries with confidence intervals of ± 1.96 SD. It cannot be clearly seen, whether surgery 2 and surgery 6 have overlapping error bars or not.

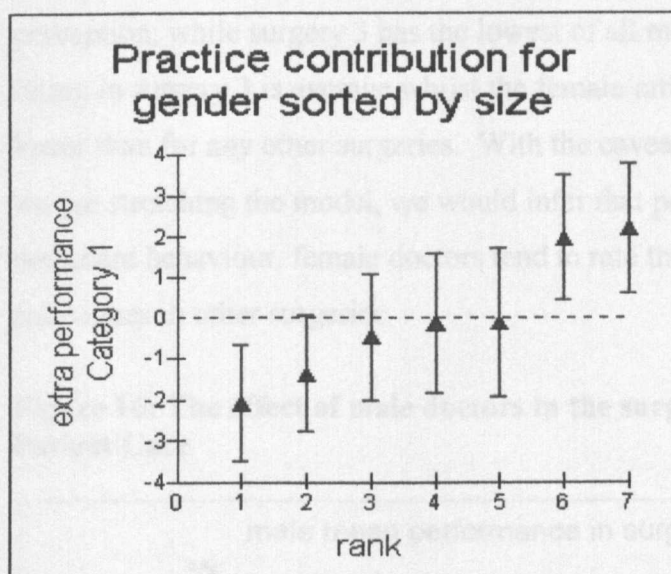
Figure 14: Gender affecting self-perception of Patient Care differs by practice



However, Snijders and Bosker refer to Goldstein and Healy (1995, p175): "... For example, testing the equality of a series of two-level residuals at the five percent significance level, requires confidence intervals that are constructed by multiplying the standard error given by 1.39 rather than the well-known five percent value of 1.96.... So, the 'comparative confidence intervals' are allowed to be narrower than the confidence intervals used for assessing single groups." Using 1.39 confidence interval separates the error bars in Figure 15, so that it now becomes obvious that surgery 2

and surgery 6 differ significantly. Thus in different practices the difference between the rating of male and female doctors for patient care is more marked.

Figure 15: Gender affecting self-perception of Patient Care differs by practice – adjusted for confidence intervals



The ranking of the surgeries cannot be explained simply by the relative number of male to female doctors in each surgery.

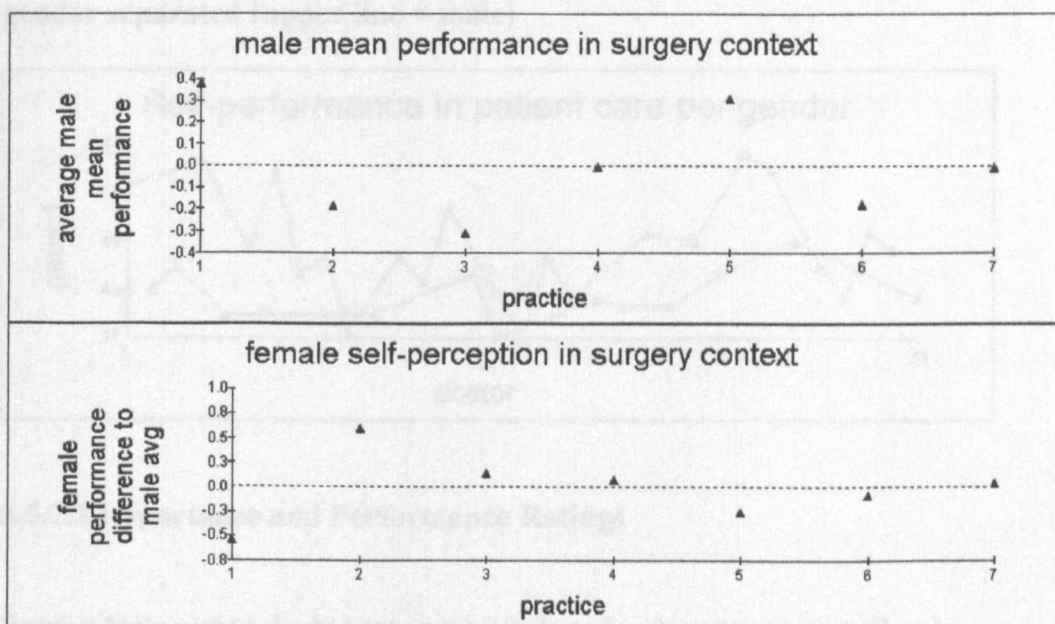
Closer inspection of the data shows that it is the combination of how well the male practitioners rated themselves AND the number of female doctors in the respective surgeries that lead to the ranking of the surgeries observed above. Female doctors evidently rate themselves lower than their male counterparts. In addition, the male doctors in the lowest ranked surgery self-assessed themselves lower than their colleagues in the highest 2 ranked surgeries, which also have a lower proportion of female doctors.

Does the tendency of male doctors to rate themselves higher vary between surgeries or is it the way that the female doctors rate themselves that varies? In order to investigate whether, across surgeries, female practitioners rate themselves differently, we calculated the group random intercepts and slopes, or more precisely the random contributions to the average slope and intercept.

Evaluating Process and Outcome in the Education of General Practitioners

In figure 16, the top plot displays the average male performance for each surgery. The lower plot depicts, for each surgery, the degree by which the female colleagues rate themselves differently than the average of their male colleagues. The plot confirms our earlier observation as to why surgeries 3 and 5 and 1 were found at the two extremes of ranking table. Surgeries 1 and 5 do indeed exhibit a "dominant" male self-perception, while surgery 3 has the lowest of all male ratings. In addition, the female rating in surgery 3 is average whilst the female rating in surgeries 1 and 5 is much lower than for any other surgeries. With the caveat that our numbers are small, and we are stretching the model, we would infer that perhaps where male doctors exhibit dominant behaviour, female doctors tend to rate themselves worse than their female colleagues in other surgeries.

Figure 16: The effect of male doctors in the surgery on female self-perception of Patient Care



Compare Figure 17, where we have clustered the self-ratings of performance in 'Patient Care' of the 33 doctors into their respective 7 surgeries, with Figure 18, where the doctors have been separated by gender, (lower line = female) so that the gender of each of the doctors in the first plot is revealed, and the aforementioned effect of surgery size and male self-perception becomes apparent.

Figure 17: Self-ratings of performance in Patient Care, clustered by surgery

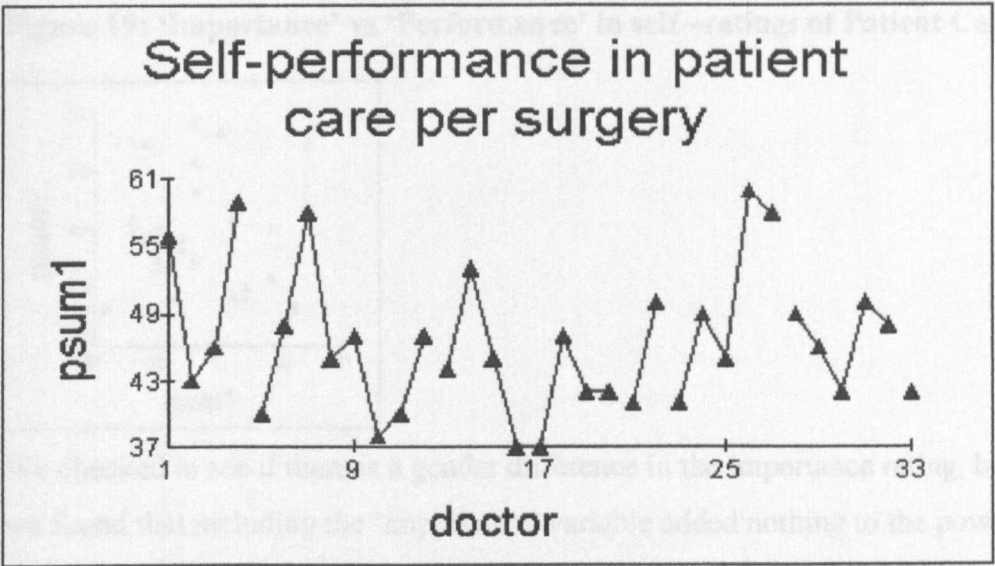
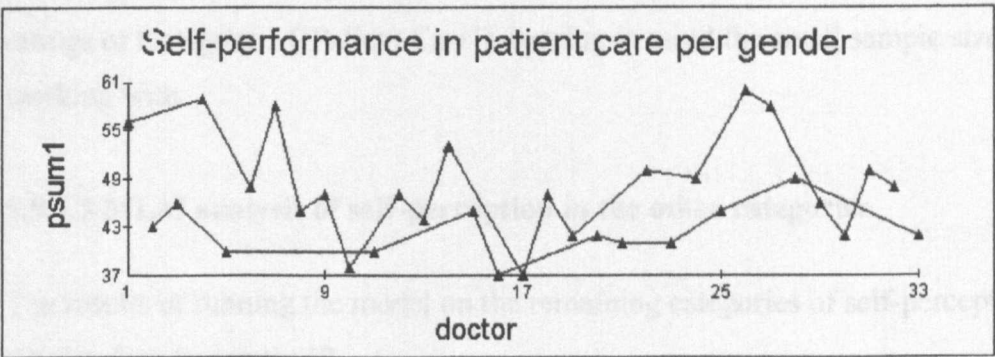


Figure 18: Self-ratings of performance in Patient Care, clustered by surgery, gender separated (upper line = male)



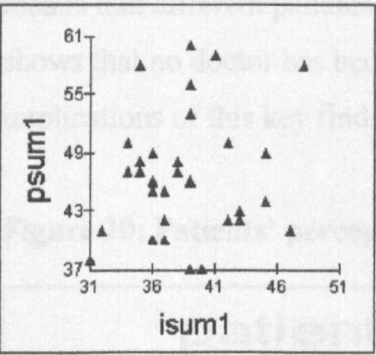
5.5.7.2 Importance and Performance Ratings

Respondents were asked to rate each attribute by importance as well as by performance. We looked to see if the model could be improved by including the factor Importance for category 1. This would show if doctors perceive themselves as performing differently for more or less important issues of the questionnaire. Hence, including the Importance factor does not explain our model better. The independence of both variables is also shown graphically in figure 19.

5.5.7.4 Patient perceptions analysed by MLM

Analysing patient information, we found no obvious differences from one surgery to

Figure 19: ‘Importance’ vs ‘Performance’ in self –ratings of Patient Care



We checked to see if there is a gender difference in the importance rating, but again we found that including the ‘importance’ variable added nothing to the power of the model to explain variance.

In our opinion, we have performed as much analysis as is reasonable on the self-ratings of Category 1 (‘Patient Care’), bearing in mind the small sample size we were working with.

5.5.7.3 MLM analysis of self-perception in the other categories

The results of running the model on the remaining categories of self-perception are tabulated in Appendix 12.

In all four categories including a gender term improved the model significantly.

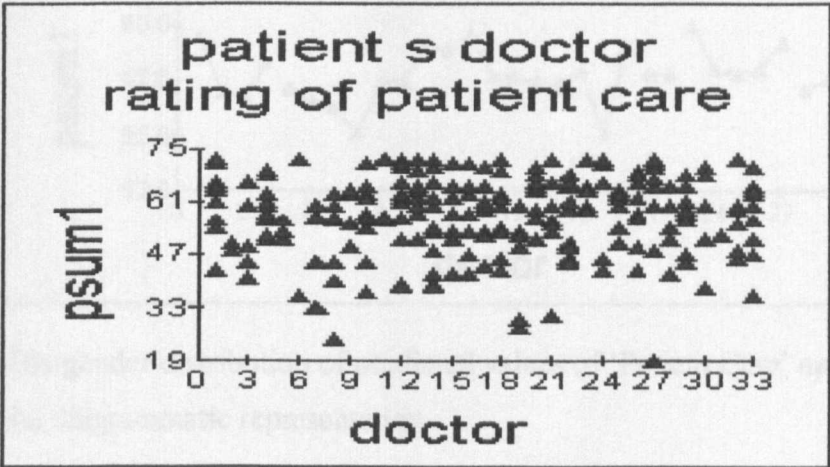
Table 5.13: Gender influence in the four categories

Category	χ^2 value (1 degree of freedom)	p-value
‘Patient Care’	9.993	0.0015714
‘Happy’	12.68	0.00036959
‘Professional’	12.368	0.00043863
‘Up-to-date’	4.284	0.038473

5.5.7.4 Patient perceptions analysed by MLM

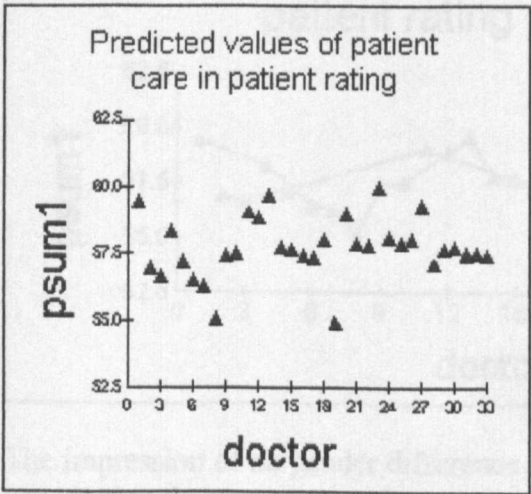
Analysing patient information, we found no obvious differences from one surgery to another. On the other hand we did find a significant variation at the patient level. This means that different patients have rated the same doctor totally differently. Figure 20 shows that no doctor has been rated consistently better or poorer than another. The implications of this key finding are explored in section 5.6.3.6.

Figure 20: Patients’ perceptions of individual doctors’ ‘Patient Care’



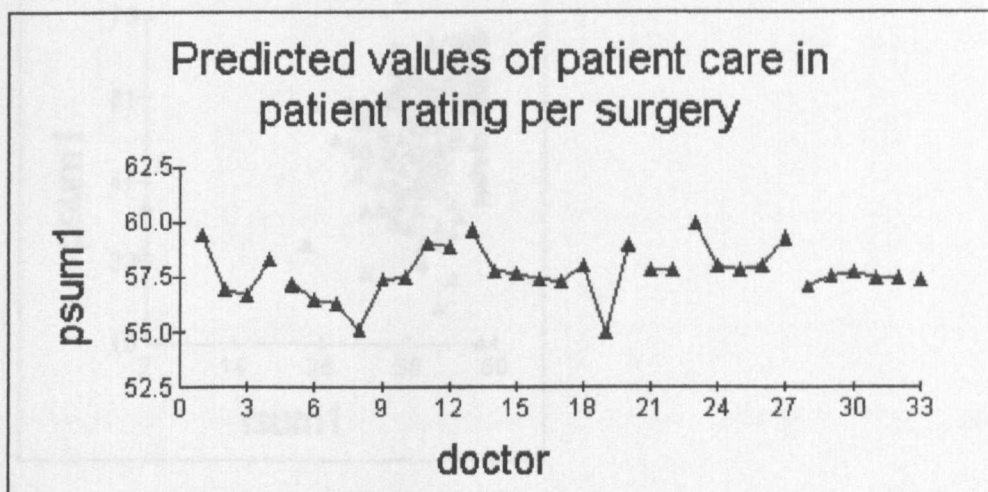
However, using a technique whereby each doctor’s rating by patients for ‘Patient Care’ is compared against ‘predicted values’ of patient rating we can identify some doctors who performed better or poorer than the average. (Figure 21)

Figure 21: Patients’ perceptions of individual doctors for ‘Patient Care’ shown against predicted values



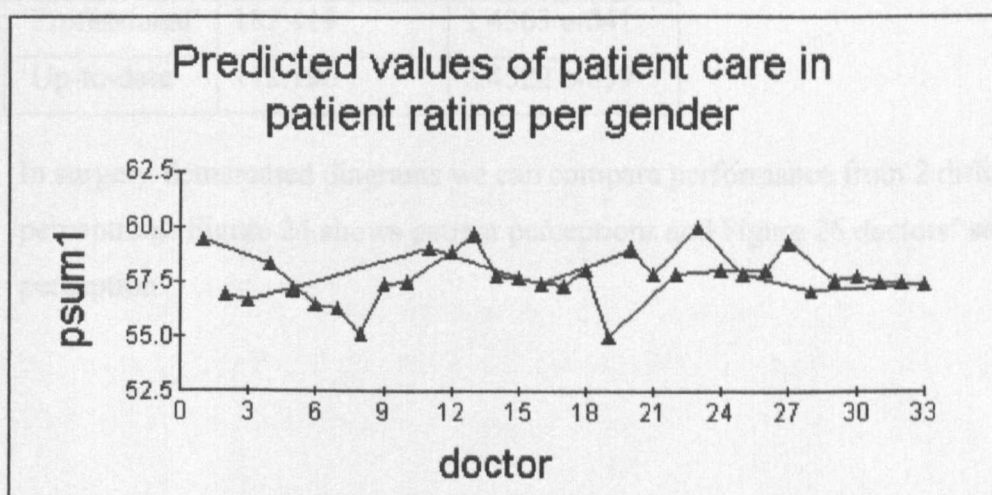
A remarkably constant pattern emerges if the series is demarcated by surgery, (Figure 22).

Figure 22:Patients' perceptions of individual doctors for 'Patient Care' clustered by surgery, and shown against predicted values



The gender distribution of predicted values of 'Patient Care' appears to be similar in the diagrammatic representation.

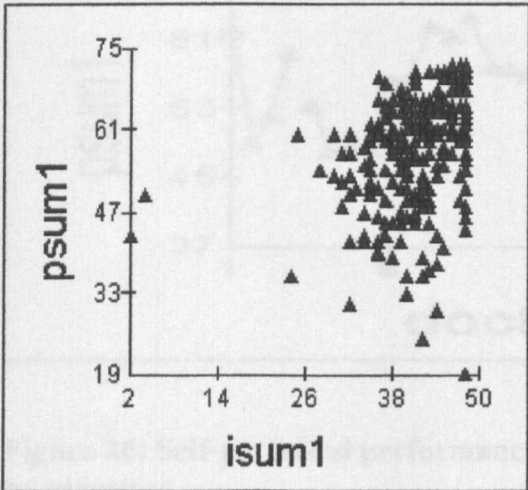
Figure 23:Patients' perceptions of individual doctors for 'Patient Care' separated by gender, (male = upper line at left), and shown against predicted values



The impression of no gender difference in Figure 23 is confirmed by the finding that introducing the gender term does not improve this model at all.

Interestingly, this time, including the factor ‘Importance’ (ISUM1) does improve the model significantly.

Figure 24: ‘Importance’ vs ‘Performance’ in patient ratings of Patient Care



What this tells us is that patients are influenced in how they rate performance of doctors in ‘Patient Care’ by how much importance they attach to the question item.

Table 5.14: ‘Importance’ influencing assessments of ‘Performance’ in the four categories

Category	χ^2 value	p-value
‘Patient Care’	121.89	2.4400 e-028
‘Happy’	168.83	1.3326 e-038
‘Professional’	182.419	1.4363 e-041
‘Up-to-date’	172.196	2.4522 e-039

In surgery-demarcated diagrams we can compare performance from 2 different perceptions: Figure 25 shows patient perceptions and Figure 26 doctors’ self-perception.

Figure 25: Patient perceived performance of doctors in 'Patient Care', demarcated by surgeries

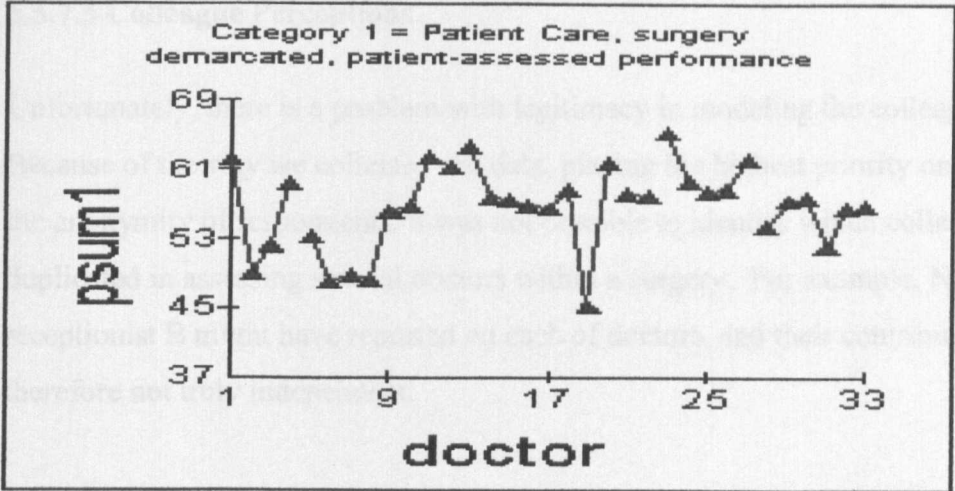
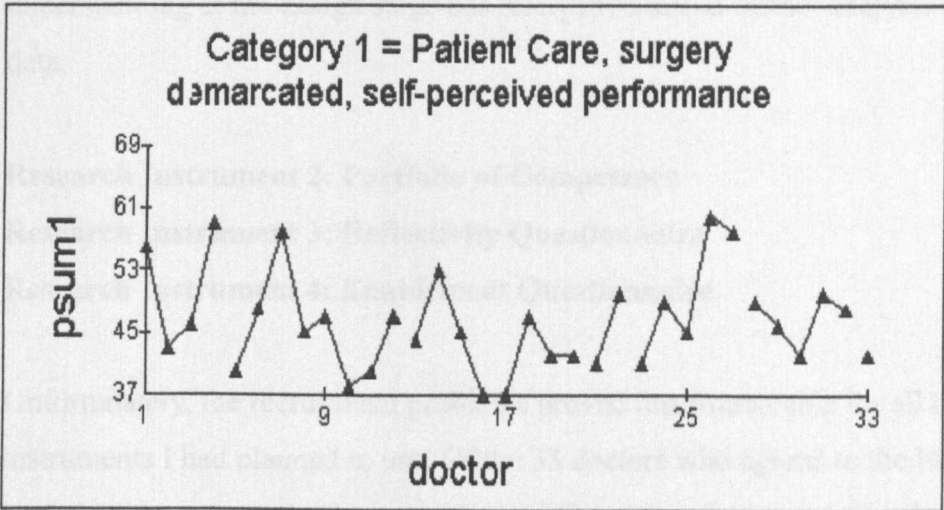


Figure 26: Self-perceived performance of doctors in 'Patient Care', demarcated by surgeries



Already in Figure 4 we have seen a slope diagram illustrating how frequently patients' perceptions differ from doctors' self-perceptions for 'Patient Care'. What Figures 25 and 26 add is a comparison of relative perceptions, surgery by surgery. Notwithstanding the similar patterns for the left-hand surgery, the absence of similarity in the other profiles suggests that there is no systematic effect of different surgeries on the patient/doctor perception gap.

5.5.7.5 Colleague Perceptions

Unfortunately, there is a problem with legitimacy in modeling the colleagues' views. Because of the way we collected our data, placing the highest priority on protecting the anonymity of respondents, it was not possible to identify which colleagues were duplicated in assessing several doctors within a surgery. For example, Nurse A or receptionist B might have reported on each of doctors, and their contributions are therefore not truly independent.

Even more importantly, doctors, who as targets are analysed at level 2, would be analysed at level 1 as 'Colleagues', and it is not permissible in a multi-level model to include the same individual in the model at two different levels. My lack of understanding at the design stage has really invalidated further analysis of colleague data.

Research Instrument 2: Portfolio of Competence

Research Instrument 3: Reflectivity Questionnaire

Research Instrument 4: Enablement Questionnaire

Unfortunately, the recruitment problems proved insurmountable for all the remaining instruments I had planned to use. Of the 33 doctors who agreed to the Insight 360 participation, none agreed to provide the information for a portfolio of competence. Several doctors kindly offered to provide the information, which they were collating for their annual appraisal, this non-standardised data would not suit our purpose. Appraisals are conducted by Primary Care Trusts,(PCTs) and until the appearance in 2002 of a standard appraisal form, there was no standardisation of format.

Only four doctors completed Reflectivity questionnaires, and this was inadequate to do more than contribute to the refinement of the questionnaire pilot work, that is ongoing.

None of the research subjects agreed to burden their patients with the enablement questionnaire. One reason for this was that many surgeries were involved in

collecting patient satisfaction data on consultations, at the behest of their PCTs. The perception-based work is therefore untriangulated by any other method.

5.6 Reflection: Phase Three

There are 3 main areas for reflection in this chapter:

- Recruitment difficulties, and the implication for future work.
- Insight 360: evaluating the instrument as a measure of quality.
- My findings' alignment with theoretical considerations around perception and multi-source feedback instruments.

5.6.1 Reflection on Phase 3 recruitment problems: the constraints on GP time

This research, as it turned out, was conducted at one of the busiest and most-pressurised times in recent General practice history. It is always difficult to persuade colleagues to take part in time-consuming research, and my research strategy depended, in part, on connecting with individual doctors' preparation for the widely anticipated compulsory revalidation which, it seemed almost certain, would include a requirement to produce data on perceived performance, in the personal portfolio of evidence. I had hoped that by providing free software and guidance through the demanding task of running 360 degree systems in the practice, doctors would feel that their research effort was mitigated by being 'ahead of the game' for revalidation. Just at the wrong time for me, the NHS appraisal system was finalised, with no requirement to provide such data, and the revalidation timetable was placed at least 5 years down the line. Short-staffed and hard pressed, even many academic practices met my requests for co-operation with 'thanks but no thanks'.

We soon realised that it was an impossible task to collect a large number of surgeries where there were doctors who had been trained in practices assessed under Phase Two, (indeed only 4 such ex-Registrars were included in our sample) and we settled for collecting as many surgeries as we could muster. In the end we managed to recruit only 33 doctors working in 7 surgeries of different sizes – enough for a few definite findings and some interesting pointers.

5.6.2 Reflections on Phase Three measurements in Insight 360

In October 2002, Pringle and colleagues (2002) published standards to apply to descriptions of quality applicable to looking for goodness in doctors. Their framework is detailed in Appendix 13, and their criteria are named in **bold** in the next paragraph.

By Pringle *et al*'s criteria, the use of Insight 360 would appear to be **Valid** and **Communicable**, and the measure is relatively easily understood. There are, however some concerns about how **Effective** the Insight 360 is, in that there may be some perverse incentives in placating patients rather than striving to make best choices in the difficult world of rationed, evidence-based medicine. The data appear reasonably consistent, but the question areas omitted by patient observers causes our instrument to fall short of Pringle *et al*'s definition of a **reliable** tool. Being perception-based, it does not meet their criteria that data should be as independent of subjective judgment as possible, but I would argue that it can still be **objective** by dint of summing the subjective. In order to be classed as **available** the data should be available quickly with minimum of extra effort and cost; and this depends on this sort of data being routinely collected for revalidation purposes. Our measure is relatively easily adjusted for important **context** effects by means of MLM. The paucity of cross-validating **comparable** (or 'gold-standard') data means that I must be cautious about how **attributable** this measure is, in that we do not know how well it reflects quality of individual doctors, nor whether it will be used appropriately in its presentation and **interpretation**.

In order to meet the parameter designated **remediable**, which requires that there be recognised, accepted, and feasible methods for influencing the measure and improving quality, we would have to demonstrate that Insight 360 feedback can lead to improvement. This has been done, using the parent instrument as a management consultancy tool. In order to be classed **repeatable**, the measure should be sensitive to improvement over time. Although Insight 360 can be run repeatedly, the data compactness is such that I very much doubt if improvement will be easily detected.

5.6.2.1 Questionnaires

The questions attracting poor levels of response from patients pose some interesting dilemmas in refining the questionnaires. Despite the fact that there are many patients who have not experienced their doctor consulting over the telephone, responding to an emergency, or responding to a complaint, is it desirable to omit such questions? Should we try to trap the responses even if few in number, because even those unable to answer about the doctor's performance on these items, rated them highly for importance. (This in itself is an argument for retaining the 'Importance' questions.)

Likewise, there were many aspects of the doctor's work upon which colleagues felt their perspective was too limited to give meaningful responses. Again, we need to reflect on how the instrument is to be used, before deciding how to modify it. If it is essential that adequate numbers of patients or colleagues answer every question, then some of these questions must be omitted in revision. If, on the other hand, it is more important to attempt to get some perspectives on even the least accessible domains of performance, then we should retain such questions. I return to this issue in the next chapter.

5.6.3 Reflection: How well do Phase Three findings accord with theory surrounding perceptions?

The key findings in my measurements were around self-perception and gender (5.5.7 & 5.5.7.1), and patient perceptions of their doctors (5.5.7.4).

5.6.3.1 What are the determinants of self-perception, other than performance in the tested domain?

The way we see ourselves reflects the "Big Five" model of personality (Bernard, Hutchison *et al.*, 1996): Ego-strength, hardiness, self-esteem, self-efficacy, and optimism. Otherwise known as the 'Health-related personality constructs', these components form 'Personal Shadow' in my conceptual model (Appendix 2.4).

5.6.3.2 Self-perceptions and gender

Perhaps the most striking finding of this phase of work has been the difference in the way that male and female doctors perceive themselves. Studies have shown that

gender differences in performance evaluations are attributable to bias (Beyer, 1990). Men hold liberal (self-serving) evaluation biases, whereas women hold conservative (self-derogatory) biases. These gender biases in self-perception interact with the nature of the task – Beyer refers to the concept of task gender-typedness. In her work, men and women were accurate in their self-perceptions on feminine tasks but women underestimated their performance on masculine tasks. The results for men evaluating performance on masculine tasks were less consistent. The implications of our findings would be that in General Practice masculine tasks might predominate. I have been unable to find data on gender-typing of General Practice, despite much talk of ‘feminisation’ of the work-force. This deserves further exploration as Beyer later discussed how females’ inaccurate self-perceptions might negatively affect achievement behaviour and curtail their participation in masculine domains, such as computer work (Beyer and Bowden, 1997).

5.6.3.3 Why is our data so compact?

The usefulness of our data is limited by the compactness: for meaningful analysis it is helpful to have widely spread data, with a lot of variance to explore. Why should the 331 patients be so similar in their perceptions on 33 doctors? Why did so little negative perception emerge? Does this truly reflect the way people in general feel about doctors? Is there a halo effect stemming from the prestige which the public ascribes to medical practitioners?

A relevant theoretical perspective is Affect Control Theory, whereby an entity’s affective meaning is measured by averaging judgments from multiple (typically 30 male and 30 female) respondents on bi-polar rating scales assessing Evaluation, Potency, and Activity (EPA) (Heise, 2002). With these tools of ‘mathematical sociology, the group at Indian have been able to compare ‘doctors’ and ‘patients’ in the American public’s profile with the profiles of specific subgroups, such as an elderly person’s group and a student group. They found that the elderly saw a doctor as somewhat less powerful than other Americans do and a patient as ‘less powerless’. The elderly people’s sentiments were interpreted as “lead(*ing*) them to want richer doctor-patient interactions presumably because doctors are so central in the lives of the elderly.” (Averett and Heise, 1988). All these studies had positive EPA ratings, in

line with our work and supporting a more universal 'halo effect' biasing patient perception of doctors. This effect may also reflect something of the awe of 'authority', as doctors are still perceived as authority figures (Rigby and Metzger, 1992).

5.6.3.4 Do perceptions still reflect meaningful dimensions of quality?

Deflection is the term used in Affect Control Theory for the difference between the fundamental sentiment and the transient feelings generated by an 'event' such as a medical consultation. Pre-existing Impressions are modified when an attribute is noted. Research suggests that the modifier (or newly noticed attribute) has about twice the impact of the 'identity', being the pre-formed impression (Heise, 2002). Thus, in evaluation, noticing that a person has a good attribute like gentleness makes the person seem more pleasant, and noticing a bad attribute like surliness makes the person seem less pleasant. A characteristic emotion is the emotion that would be experienced were an individual's identity confirmed perfectly.

According to Affect Control Theory (Heise, 2002), individuals seek experiences that confirm fundamental affective meanings. We read into social interactions the behaviours which are concordant with our fundamental affective meanings, and are inclined to enact such behaviours, or ask others to enact them, in order to obtain confirming experiences. Heise (2002), taking an example of doctor and a patient, uses a simulation model of Affect Control Theory to show how emergent meanings of doctor and patient combine with sentiments about doctors and patients to produce the structural emotions of the relationship - the doctor feeling secure and compassionate, and the patient feeling at-ease and grateful.

What I take from this is that the overtones of the Doctor-patient role stigmatisation may be so strong as to mute the individual judgements of patients about doctors, and may account for some of the data compactness.

5.6.3.5 What are the other Biases in Person Perception?

The potential biases in perceptions of others have been categorised by (George and Jones, 2002), and are presented in Table 5.13.

Table 5.15: Biases and Problems in Person Perception (after George, 2002)

Primacy Effects	First impressions: perceivers have been shown to be inordinately affected by the initial pieces of information about a target, and this will affect all subsequent evaluation.
Contrast Effects	The perceiver's perceptions of others distort the perceiver's perception of a target. Thus an 'average' target may 'shine' in comparison to a 'mediocre' group
Halo Effect	The perceiver's general impression of a target distorts his or her perception of the target on specific dimensions.
Similar-to-me Effects	People perceive others who are similar to themselves more positively than they perceive those who are dissimilar.
Harshness, Leniency, & Average Tendency	Some perceivers tend to be overly harsh in their perceptions, some overly lenient. Others view most targets as being about average.
Knowledge-of-Predictor Bias	Knowing how a target stands on a predictor of performance influences perceptions of the target.

The way we attribute personality traits, motives and abilities to other people, is the subject of Attribution Theory (Hilton and Slugoski, 1986). Anderson proposes that our judgments of other people are weighted averages of the information we have about them (Anderson (1981) cited in Schlottmann and Anderson, 1995). In other words, in reaching decisions about other peoples' personality, abilities and so on, we tend to take everything we know about them, giving some information a greater weight, and then produce an average of the information we have. Negative information will generally be given more weight than positive information.

5.6.3.6 Consistency of Patient Perceptions

MLM (see section 5.5.7.4 and Appendices 11 and 12) shows that most variation in scores is between patients, rather than between doctors or practices. What this means is, that individuals vary in their perception more than any consistent difference in the quality (assessed by perception) of doctors or practices. This finding undermines the

Evaluating Process and Outcome in the Education of General Practitioners

use of patient perception as a tool in routine quality assessment of GPs. Given that all patient opinion tends to lie within a relatively narrow band, for any one doctor, we can expect to find a wide scatter of patient opinion, whatever the practice context.

What it does, however, open up is the use of patient perceptions in the possible identification of under-performing doctors. Thus, if the small amount of variability in perceived performance is such that the average doctor does not stand out, then it is worth examining outliers for consistency of patient perception, as a high degree of consistency in low rating is truly remarkable. We observed no obvious outliers in 331 patient observations on 33 doctors, so I cannot produce evidence about an individual outlier. I can state however, that finding an outlier in patient reporting, particularly on Patient Care, would make me prick up my ears.

5.6.3.7 Speculating about individual perceptions

What can we read into the perceptions of doctor 15 (see Table 5.11). Possible explanations include:

- The doctor is unpopular with colleagues
- The doctor under-performs in a number of areas and colleagues are best placed to recognise this
- Patients can recognise the doctor under-performing in the area of 'Professional' performance, but a 'Positive regard' halo effect, prevents them recognising low performance in respect of 'Patient Care'
- The doctor is able to recognise his/her own underperformance in respect of the 'Professional' performance domains, but for other areas scores self only marginally below average.

Or again, speculating on another outlier, doctor 22 (see Table 5.12).

- Is 'Professional' the doctor's blind-spot, where a distorted self-perception impedes a much-needed improvement in performance?
- This doctor self rates above average in all domains, whereas colleagues rate him/her marginally or significantly below average in all domains – does this signify an arrogance that gets in the way of team-working.

Patients rate this doctor close to the mean in terms of “Patient Care” and not significantly lower in ‘Professional’ but he/she does less well in patients’ eyes on ‘Happy’ and significantly better on ‘Up-to-date’. (See the slope diagram outlier). Are we looking at a miserable doctor, who doesn’t get on very well with colleagues, but keeps up-to date?

The general consistency of our results within doctors, suggests that perceptions are accessing information about performance. However, the above discussion raises some important points, which need to be picked up in the next chapter.

Speculating serves to remind us that this sort of questionnaire-based perception tool cannot in isolation confirm underperformance or offer in-depth analysis of performance or competence. At best, it can achieve one of two objectives (Table 5.16).

Table 5.16: Legitimate uses for perception-based multi-source performance data

<ol style="list-style-type: none">1. To illuminate differences in perception in order to assist the individual doctor’s process of reflection.2. In carefully analysed group comparison, to highlight individuals whose performance may merit further examination by other means, including direct observation and outcome measures.

CHAPTER 6
REFLECTIONS ON THE RESEARCH PROJECT
What has been achieved? What Remains to be done?

6.1 The original intentions reviewed

The original objective of this work was to develop a methodology that could robustly relate the process, by which a GP Trainer educates a Registrar, to the outcome, in terms of the quality of the trained doctor in later practice. Clearly this objective has not been realised, in that, although the methods I used may be appropriate for the task, I have so far failed to validate a method for relating process to outcome.

The major problem was the difficulty with recruiting doctors to a complex and demanding research project at a time when GPs felt in crisis. As a result of the poor recruitment, there is a distinct lack of data on outcome measures. I have had to subject a small quantity of data to intensive analysis in order to try to define useful analytical processes for the future.

6.1.1 What has been achieved?

The significant contributions made in this research are:

1. The definition of categories and dimensions by which to describe the educational process in training practices.
2. The accumulation and documentation of useful experience in the assessment of educational processes in training practices.
3. Shedding light on practical and theoretical aspects of the use of perceptions and multi-source feedback data in the assessment of quality of performance in General Practice. This will inform further debate.
4. Pioneering the use of multi-level models of multi-source data comprising self-perceptions by doctors and parallel perceptions by patients

I propose to look at each of these components in more detail.

6.2 Categories and Dimensions of Educational Behaviours

In Chapter One, I stated: It is my contention that the quality framework for defining practice, for recording perceptions, and for assessing and improving practice is essentially similar across both educational and clinical domains of practice.

Evaluating Process and Outcome in the Education of General Practitioners

In Chapter Two, I went on to contend that a large part of the reason for the paucity of evidence linking process to outcome in GP Registrar education is attributable to the difficulty of agreeing meaningful categorisation of process and meaningful measures of outcome.

Phase One of my work involved the synthesis of grounded theoretical categories and dimensions to describe the educational behaviours in practice. Since they were published, these categories have received considerable attention, certainly locally, and have proved robust. They are in widespread use in Oxford Deanery on training courses for Trainers and Registrars, and visiting teams have become more adept at assessing training in practice, using these categories as a framework to give feedback to Trainers.

6.2.1 Scope for refinement of Categories of Educational Behaviours

The evidence from Table 4.6 is that the categories that visiting teams found most difficult to assess were Category 5 (Learning Cycles) and Category 6 (Family practice in context). In the case of Category 5, the teams accumulated much conflicting evidence. This was also a problem for Category 6, but here there was an equal and opposite problem of a total failure to collect any evidence.

In discussion with Team-leaders, a similar problem arose on 2 different occasions with Category 5. On each occasion, the Registrar had denied any teaching about change management, but the Trainer (interviewed later) was able to give cogent examples that such teaching has taken place. It may be that there is some very effective learning about change management that is so subtle that the learner does not realise it has taken place – an example of osmotic learning (Claxton, 1997). On the other hand, Trainers could be deluding themselves.

Now that 8-part audit (involving completing the cycle) has been introduced as summative assessment, it may be that we could omit Category 5 from any assessment process, on the grounds that there is at least now some drive from the summative assessment process to ingrain cycle completing behaviours in change management. For the time being, I would prefer to see Trainers being assessed under Category 5, as I think it is quite a difficult educational behaviour to master.

Evaluating Process and Outcome in the Education of General Practitioners

For Category 6, it is also important, yet difficult, to ascertain if GP Trainers and training practices are adequately contextualising the learning. I believe this should remain an assessed Category, if only to prompt change behaviour in those Trainers who are poor contextualisers.

6.3 The assessment of educational behaviours in GP training practices

I am encouraged by the experience of Team-leaders that having a schema helped them to engage in the assessment of educational behaviours more than previously. There has always been a tendency to measure the measurable, implying a tendency to look at structure rather than function on peer-review visits, and it appears that our work has helped to restore some balance, with more of an emphasis on educational process being encouraged.

Significantly, it has been the GP Trainers welcoming the feedback, which has helped to embed this process in Oxford Deanery. The face validity of the constructs with which appraisers and appraisees have engaged has been the strength of this innovation.

6.3.1 Overcoming problems of documenting behaviours on visits

What has not worked well has been the documentation of educational behaviours in the course of peer assessment visits. A crowded visit has not really left time for the recording of the numerous fragments of evidence needed to make judgements about prevailing behaviours.

Many deaneries, including Oxford-PGMDE, are placing more emphasis on the educator self-appraising against criteria, with the visiting team validating the self-appraisal. It seems to me that this offers a way forward. I have devised a method of adapting our appraisal, based on the Categories and Dimensions of Appendix 2, into a form of self-appraisal, (Appendix 16), which can be validated by visiting teams. The ten-point scale allows Trainers to track movement as their behaviour changes over the years. It is also highly beneficial to the process of future research, generating fine-scale data ideal for multi-level modelling. It is very exciting that the educational

Evaluating Process and Outcome in the Education of General Practitioners

strategy meeting of Oxford-PGMDE endorsed the adoption of this process, which will be introduced in 2003. Because there is so much that is new about the new process of self-evaluating against criteria, the advisory group agreed that there should be a staged introduction of the changes.

Also in the course of 2003, I hope to take forward the research with Dr Simon Street, looking at the process by which visiting teams obtain evidence to validate Trainers' self-appraisal.

6.3.2 Changing educators' behaviours

As discussed in Chapter 4, the peer assessment process means that doctors who will themselves be assessed, are learning the tools of assessment, and becoming familiar with preferred behaviours. I prioritised changing educator behaviour above 'pure' research considerations, and I continue to do so. The progression towards self-assessment is a further step in embedding change (Holmboe and Hawkins, 1998).

6.3.3 The Clinician Educator as Role Model

Trainer assessment processes, certainly in our region, are moving away from looking at quality of the Trainer and training practice in the delivery of care to patients. These considerations are no less important than before, but they are increasingly the remit of Clinical Governance processes, and Deaneries are resisting duplication of effort in assessing the same dimensions twice. Thus the emphasis is increasingly on looking at the quality of the Trainer and the training practice in delivering education. Deaneries will increasingly accept Clinical Governance data as evidence of quality of care provision.

There is, of course overlap between clinical practice and education, and this comes in the area of modelling. The role model that the clinician presents to the learner, is seen to be as important by my work with GP Registrars as it has previously been demonstrated in hospital residents (Wright, 1996; Wright, Kern *et al.*, 1998; Maudsley, 2001).

6.3.4 Committees sitting in judgement

In Chapter 4, much of the focus was on the process for decision-making in the Trainer Selection Committee. Although all committee members see the same report from the Team-leader, the process of committee discussion reveals differing perspectives. Our recommendation that prior knowledge should feed into the process of collecting evidence, and not that of making judgements, is being acted upon.

Other changes are happening. The process of moving more towards trainer self-assessment with validation by visiting peers, is well under way, and the whole assessment documentation has been redesigned for this purpose. The framework for self-assessment has been built around the defined criteria.

Self-assessment on educational behaviours will fit into this pattern, with each prospective trainer having his or her own evidence on prevailing educational behaviours, according to my framework, validated by the visiting team.

The committee will still have to make judgements. Evidence, even in a defined framework around defined criteria, (whether self assessment or validating external observations) is never going to be entirely clear-cut in respect of human behaviour. The judgements will become easier to define and defend with clear frameworks. Our evidence shows more consistent judgements from the core committee members. This raises the interesting question of whether the advantages of including many stakeholders in such a committee outweigh the disadvantages that lack of experience in the judgement-making process leads to less consistent decision-making.

6.4 Perceptions and multi-source feedback data

Engaging with this research, I initially tried to find outcomes in terms of quality of performance that were objective and measurable. It did not take long to realize that there are no short-cuts in the definition of quality markers, and that the state of the art is that most patient outcome data requires such careful qualitative interpretation, that

Evaluating Process and Outcome in the Education of General Practitioners

it is not yet usable for my purposes. I was rapidly thrown back on the use of perceptions, and to relying on the precept, “I know a good doctor if I see one”.

I found it helpful to engage, through conceptual modeling, with some of the theoretical considerations around the use of perceptions in appraisal and assessment, posing a number of questions to myself in Table 1. Some of these questions I was able to answer, at least in part, by reference to the literature, others became clearer as a result of our work with 360 degree perceptions.

Synthesising my own observations with the literature, I am convinced of the value of multi-source data. There is enormous value to the individual doctor in triangulating a self-perception against those of colleagues and patients. There is broad agreement on the number of observations needed, and the Insight 360 targets of 15 patients and 15 colleagues per doctor are endorsed both by the literature and by our experience, as is the minimum requirement of 4 observations per set before any report is compiled, in order to preserve respondent confidentiality.

For any one target, I do not think the diversity of opinion within a reporting set is very illuminating, as there is such wide diversity of patient opinion that the mean or predictive values are more helpful. For each observer, there will have been a process of belief revision (Schlottmann and Anderson, 1995) as experiences of the doctor mount up. There is probably a sequential anchoring-and-adjustment process as current beliefs are adjusted in the light of new information (Hogarth, 1992). Thus, as I see it, even over-riding the requirements of confidentiality to feedback individual observer perceptions to the target GP, would only provide data from a frozen instant in a changing continuum. The amalgamation of perceptions from different individuals at different stages in their anchoring-and-adjustment processes is probably ultimately more meaningful. Seen this way, the effect of recall bias is mitigated by a process of ‘weighted averaging’ of perceptions (Anderson, 1981).

This research is not able to shed light on how colleagues’ perceptions of the doctor as a care-giver influence perceptions of the doctor as an educator, but this is a consideration which I will take into account in designing further research.

6.4.1 Formative Use of Perceptions and feedback

Used formatively, does it matter if perceptions accurately reflect the reality of the doctor-patient experience? Whose reality are we dealing with anyway? It is the perceptions that are relevant to the doctor's reflection for personal growth, and if constructs of reality differ, that may be helpfully exposed. This, of course, presupposes a degree of robustness and professionalism on the part of the doctor, in order to take account of the perceptions of others without unduly unbalancing self-perception.

Of course the above considerations only apply in the formative domain of doctors self-appraising and reflecting for the purpose of personal growth. Comparisons between doctors are another matter. I had thought initially that there might be scope for groups of doctors to compare perceived strengths and weaknesses, for example, within Trainer groups. We found, however that the 'compactness' of our data was such that such comparisons are of little value. Most patients rate doctors so highly that meaningful separation on the sort of scales we used is very difficult.

6.4.2 Potential for 360 degree perspectives feeding into the revalidation process

The corollary of data compactness is, however, that outliers stand out. I believe that Insight 360 data might have some uses for the purposes of revalidation in looking for outliers. What is essential is that no facile interpretation is placed on the evidence of a doctor outlying. I have demonstrated the extent of speculation that is possible around the finding of outliers (Doctor 15 and Doctor 22) in Chapter 5. If Insight 360 is to be used in such a fashion, it must be only as a stimulus to further in-depth examination of the reasons for outlying behaviour.

A further very important caveat from our work is that the analysis of self-perceptions in doctoring necessitates having gender-separated data, as there is such a strong effect of gender on self perception.

6.5 Method of analysis

It is doubtful if even quite important findings, like the polarisation by gender of self-perceptions of performance, would have emerged were it not for the use of MLM.

We were able to run enough models on our data to be convinced of the appropriacy of this method for relating process to outcome in medical education.

There is no mechanism in MLM for performing ‘power calculations’ for determining sample size in advance of a study. As yet, the process remains one of trial and error. We had examples, as when we tried to introduce a random slope model to see whether the tendency of doctors to rate themselves higher varies between surgeries, when we found the model to be working at the limits of its capacity with inconsistent results on serial runs, meaning that we could not make meaningful assertions on this point. Our ‘feel’ for the robustness of the models we tried to run, gives us confidence that meaningful modeling of quality data would be practical across most PCTs. Using multi-source data, at least ten practices contributing data from three or more GPs would, I believe, allow modeling across the levels of individual and practice to look at variables such as gender and years of experience.

6.5.1 Will it ever be possible to relate educational process to outcome?

Even such ‘woolly’ statements as the one above are an advance on the previous uncharted situation. I would go further, and state my confidence that the modeling techniques we have used are likely to prove capable of relating educational process to outcome across a Postgraduate Deanery. If we had educational process data on 50 or more training practices, there is little doubt in my mind that multilevel modeling would be able to shed light on differences in training impacting on later doctoring. That ‘feel’ for how the models have been running is as close to a ‘power calculation’ as anyone can get at the present time! Our experience of modeling and the theoretical considerations around statistical process detailed in Chapter 5, leads me to assert that multilevel modeling is the most appropriate tool for such analyses, if not the only one capable of performing such functions.

6.6 The way forward – further research

This work has proved to be an extremely valuable pilot in the difficult field of relating process to outcome in medical education. I can claim to have made considerable strides in developing the necessary methodology, but there is a long way to go before causal relationships can be established. The categories and dimensions should be developed as a self-appraisal tool, as in Appendix 16. Insight 360 remains a useful part of the outcome methodology, and I am in discussion with the General Medical Council about the intention to use a form of 360 degree appraisal in future revalidation work. It is crucial to meaningful research that there should be broad application of the outcome measure across a large population of doctors, and only some form of compulsion (inherent in the revalidation process) is likely to ensure 'recruitment' of a large enough and representative enough population. I foresee the collation of validated self-appraisal of educational behaviours by Trainers across a Deanery being related to quality markers from revalidation, by multilevel models, such as those we have evolved. In about 5 years' time it should be possible to validate helpful educational behaviours.

Evaluating Process and Outcome in the Education of General Practitioners

- Aitkin, M., D. Anderson, et al. (1981). "Statistical Modelling on Teaching Styles (with discussion)." J Royal Statist. Soc A144: 148-161.
- Anderson, L. and R. Dedrick (1990). "Development of the Trust in Physician scale: a measure to assess interpersonal trust in patient-physician relationships." Psychological Reports 67(3 Part 2): 1091-100.
- Anderson, L., W. Rakowski, et al. (1988). "Satisfaction with clinical encounters among residents and geriatric patients." Journal of Medical Education. 63: 447-55.
- Anderson, N. H. (1981). Foundations of information integration theory. New York, Academic Press.
- Anyon, P. (1987). An evaluation of the New Zealand family medicine training programme. Wellington, Monograph, RNZCGP.
- Ashley, E. A. (2000). "Medical education - beyond tomorrow? The new doctor - Asclepiad or Logiatros?" Medical Education 34(6): 455-9.
- Averett, C. and D. Heise (1988). Modified social identities: Amalgamations, Attributions, and Emotions. Analyzing Social Interaction: Advances in Affect Control Theory. L. Smith-Lovin and D. Heise. New York, Gordon and Breach Science Publishers.
- Baker, R. (1997). "Pragmatic model of patient satisfaction in general practice: progress towards a theory." Quality in Health Care 6: 201-204.
- Barrington, D. and C. Silagy (1996). "Skills training in medical education: what skills and when should they be introduced?" Education for General Practice 7: 16-22.
- Bashook, P. and J. Parboosingh (1998). "Continuing medical education: Recertification and the maintenance of competence." BMJ 316: 545-548.
- Benner, P. (1984). From Novice to Expert: Excellence and Power in Clinical Nursing Practice. Menlo Park, Calif., Addison-Wesley.
- Bernard, L. C., S. Hutchison, et al. (1996). "Ego-strength, hardiness, self-esteem, self-efficacy, optimism, and maladjustment: Health-related personality constructs and the "Big Five" model of personality." Assessment 3((2)): 115-131.
- Bernstein, B. (1971). On the classification and framing of educational knowledge. Knowledge and Control. M. Young. London, Collier-Macmillan.
- Beyer, S. (1990). "Gender differences in the accuracy of self-evaluations of performance." Journal of Personality and Social Psychology 59: 960-970.
- Beyer, S. and E. M. Bowden (1997). "Gender differences in self-perceptions: Convergent evidence from three measures of accuracy and bias." Personality and Social Psychology Bulletin, 23: 157-172.
- Black, G. and L. Macdonald (2000). "Practice-based learning: an alternative approach to continuing medical education in primary care." Education for General Practice 11: 397-404.
- Bligh, J. G. (1992). "Independent learning among general practice trainees: an initial survey." Medical Education(26): 497-502.
- Bligh, J. and P. Slade (1996). "A questionnaire examining learning in general practice." Medical Education 30(1): 65-70.
- Boaden, N. and J. Bligh (1999). Community-Based Medical Education. London, Arnold.
- Brahams, D. (1982). "Assessment of trainers by trainees." Practitioner 226(1372): 1652-3.
- Brigley, S. (2002). Developing reflective clinical practice. Effective Learning and Teaching in Medical, Dental and Veterinary Education. J. Sweet, S. Huttly and I. Taylor. London, Kogan Page.

Evaluating Process and Outcome in the Education of General Practitioners

- Buchler, J. (1961). The Concept of Method. New York, Columbia University Press.
- Burrows, P., N. Gravestock, et al. (2001). "Assessing clinical competence and revalidation of clinicians." BMJ **322**: 1600-1601.
- Bush, T. (1983) "The role of vice-principal in Further and Higher Education." Journal of Further and Higher Education **7**: 16-23.
- Calman, K. (1994). "The profession of Medicine." British Medical Journal **309**: 1140-3.
- Campbell, S. M., J. A. Cantrill, et al. (2000). "Prescribing indicators for UK general practice: Delphi consultation study." BMJ **321**: 425-428.
- Charlton, R., S. Field, et al. (2000). "The effect of the general practice registrar year on perceived skills in palliative care in the West Midlands." Medical Education **34**: 928-35.
- Cicourel, A. V. (1976). Social Organisation and juvenile justice. London, Heinemann.
- Claxton, G. (1997). Hare Brain, Tortoise Mind. London, Fourth Estate.
- Cohen, L. and L. Mannion (1994). Research Methods in Education. London, Routledge
- Connors, G. and T. Munro (2001). "360-degree physician evaluations." Healthcare Executive: 58-59.
- Cornbleth, C. (1991). Research on Context, Research in Context. Handbook of Research on Social Studies Teaching and Learning. J. E. Shaver. New York, MacMillan: 265-275.
- Costa, A. and B. Kallick (1993). "Through the lense of a critical friend." Educational Leadership **51**: 49-51.
- Coulter, A. and G. Elwyn (2002). "What do patients want from high-quality general practice and how do we involve them in improvement." British Journal of General Practice, **52 (suppl)**: S22-S26.
- Crawley, H. S. and J. B. Levin (1990). "Training for general practice: result of a survey into the general practitioner trainee scheme." British Medical Journal(300): 911-915.
- Cunnington, J., E. Hanna, et al. (1997). "Defensible assessment of the competency of the practicing physician." Academic Medicine **72**.
- Davis, D., M. Thomson, et al. (1992). "Evidence for the effectiveness of CME. A review of 50 randomized controlled trials." JAMA **268**: 1111-7.
- Davis, D., M. Thomson, et al. (1995). "Changing physician performance. A systematic review of the effect of continuing medical education strategies." JAMA **274**: 700-705.
- Davis, P. (2001). "The interface between primary and secondary care." Journal of the Royal Society of Medicine **94 Suppl 39**: 46-9.
- Difford, F. and R. Hughes (1991). "Experience of using rating scales for the assessment of vocational trainees in general practice." British Journal of General Practice, **41**: 360-4.
- Dixon, J., P. Holland, et al. (1998). "Primary care: core values developing primary care: gatekeeping, commissioning, and managed care." BMJ **317**: 125-8.
- DoH (2002). NHS Appraisal :Appraisal for General Practitioners working in the NHS. London, Department of Health.
- Dolmans, D., I. Wolhagen, et al. (1996). "Long-term stability of tutor performance. Source." Academic Medicine, **71**: 1344-7.
- Dreyfus, H. L. and S. E. Dreyfus (1986). Mind over Machine: The Power of Human Intuition and Expertise in the Era of the Computer. Oxford, Basil Blackwell.
- Duncan, P. (1994). "Evaluation of the GPVTP." The New Zealand Family Physician(21): 186-8.

Evaluating Process and Outcome in the Education of General Practitioners

- Epstein, R. and E. Hundert (2002). "Defining and Assessing Professional Competence." JAMA **287**: 226-235.
- Erasmus, R., P. Coetzer, et al. (1977). "Medical problem-solving: a five stage approach in primary medical care." Medical Education **31**.
- Eraut, M. (1994). Developing Professional Knowledge and Competence. London, Flamer Press.
- Fidler, B., P. Earley, et al. (1998). "Teacher Gradings and OfSTED Inspection: help or hindrance as a management tool?" School Leadership and Management **18**(2): 257-270.
- Finucane, P., L. Allery, et al. (1992). "Teachers at a British medical school." Medical Teacher, **14**: 275-82.
- Frankel, S., J. Sterne, et al. (2000). "Mortality variations as a measure of general practitioner performance: implications of the Shipman case." BMJ **320**: 489.
- Freeman, F. and P. S. Byrne (1976). The assessment of vocational training for general practice. Reports from General Practice. London, Royal College of General Practitioners.
- Freer, C. B. and M. E. Reid (1978). "A survey of past trainees. Vocational training forum." Update(16): 1461-1468.
- Freiberg, H., H. Waxman, et al. (1987). "Enriching feedback to student teachers through small group discussion." Teacher Education Quarterly **14**: 71-82.
- George, J. and G. Jones (2002). Understanding and Managing Organizational Behavior. New Jersey, Prentice Hall.
- Gerteis, M., S. Edgar-Levitan, et al. (1993). Through the patient's eyes: understanding and promoting patient-centred care. San Francisco, Jossey Bass.
- GMC (2001). Good Medical Practice (Third Edition). London, General Medical Council.
- GMC and DoH (2002). The guide to appraisal and revalidation-The Toolkit.
- Goldstein, H. (1995). Multilevel Statistical Models. London, Edward Arnold.
- Goldstein, H. and M. Healy (1995). "The graphical presentation of a collection of means." Journal of the Royal Statistical Society, **158**(Ser. A).
- Goldstein, H., J. Rasbash, et al. (1993). "A multilevel analysis of school examination results." Oxford Review of Education **19**: 425-433.
- Gordon, M. (1991). "A review of the validity and accuracy of self-assessments in health professions training." Acad Med **66**: 762-9.
- GPC (2001). National Survey of GP Opinion. London, British Medical Association.
- Gray, J. (1996). "Global rating scales in residency education." Academic Medicine **71**: S55-63.
- Griffin, E., C. Sanders, et al. (2000). "A computerised 360° feedback tool for personal and organisational development in general practice." Health Informatics Journal, **6**:251-9
- Grol, R. (1997). "Beliefs and evidence in changing clinical practice." BMJ **315**: 418-421.
- Grol, R., H. Mokkink, et al. (1989). "Effects of the vocational training of general practice consultation skills and medical performance." Medical Education **23**(6): 512-21.
- Grow, G. (1991). "Teaching learners to be self-directed." Adult Education Quarterly **41**: 125-149.
- Hall, W., C. Violato, et al. (1999). "Assessment of physician performance in Alberta." CMAJ **161**: 52-57.

Evaluating Process and Outcome in the Education of General Practitioners

- Harre, R (1970), "The Principles of Scientific Thinking", University of Chicago Press, Chicago
- Harre, R. (1985). The philosophy of science (2nd ed.). Oxford: Oxford University Press.
- Hayden, J. and J. Adams (2000). "Assessment of competence and performance at interview." BMJ **321**: 449.
- Heise, D. (2002). Understanding Social Interaction With Affect Control Theory. New Directions in Sociological Theory. G. Berger and M. Zelditch. Boulder CO, Rowman and Littlefield.
- Heron, J. (1999). The Complete Facilitator's Handbook. London, Kogan Page.
- Hilton, D. and B. Slugoski (1986). "Knowledge-based causal attribution: The abnormal conditions focus model." Psychological Review **93**: 75-88.
- Hilton, S. (1981). "Looking back on a trainee year." Practitioner **225**(1354): 443-4.
- Hindmarsh, J. H., G. D. Coster, et al. (1998). "Are vocationally trained general practitioners better GPs? A review of research designs and outcomes." Medical Education **32**(3): 244-54.
- Hodkinson, P. and M. Issitt (1995). The Challenge of Competence. London, Cassell.
- Hogarth, R.M. & Einhorn, H.J. (1992). "Order effects in belief updating: The belief-adjustment model." Cognitive Psychology **24**: 1-55.
- Holmboe, E. and R. Hawkins (1998). "Methods for Evaluating the Clinical Competence of Residents in Internal Medicine: A Review." Annals of Internal Medicine **129**: 42-48.
- Howie, J., D. Heaney, et al. (2000). "Developing a 'consultation quality index' (CQI) for use in general practice." Family Practice **17**: 455-461.
- Howie, J., D. Heaney, et al. (1999). "Quality at general practice consultations: cross-sectional survey." BMJ **319**: 736- 743.
- Hox, J. (1995). Applied Multilevel Analysis. Amsterdam, TT-Publikaties.
- Hurwitz, B. and A. Vass (2002). "What is a good doctor and how can you make one." BMJ **325**: 667-8.
- Hyland, T. (1994). Competence, Education and NVOs: dissenting perspectives. London, Cassell
- Irby, D. (1990). "Shifting paradigms of research in medical education." Acad Med **65**: 622-3.
- Irvine, D. (1997). "The performance of doctors. I: Professionalism and self regulation in a changing world." BMJ **314**: 1540-1542.
- Irvine, D. and L. Donaldson (1993). "Quality and standards in health care." Proceedings of the Royal Society of Edinburgh **101B**: 1-30.
- Janis, I. L. (1982). Groupthink: Psychological studies of policy decisions and fiascoes. Boston, Houghton Mifflin.
- JCPTGP (2003). Joint Committee on Postgraduate Training for General Practice. <http://www.jcptgp.org.uk> accessed 11 January, 2003
- Johnson, N. and J. Hasler (1997). "Content validity of a trainers' report: Summative assessment in general practice." Medical Education **31**(4): 287-92.
- Johnson, N., J. C. Hasler, et al. (1997). "The role of a practice manager in training practice assessment visits." Education for General Practice **8**(2): 128-34.
- Jones, R. and J. Froom (1994). "Faculty and administration views of problems in faculty evaluation." Academic Medicine. **69**: 476-83.
- Joyce, B. and B. Showers (1980). "Improving in-service training: the messages of research." Educational Leadership **37**: 379-385.

Evaluating Process and Outcome in the Education of General Practitioners

- Kagan, N., & Kagan, H. (1991). Interpersonal process recall. In P. Dowrick (Ed.), Practical guide to using video in the behavioral sciences (pp. 221-230). New York: John Wiley & Sons.
- Katon, W., C. Rutter, et al. (2000). "Are there detectable differences in quality of care or outcome of depression across primary care providers?" Medical Care 38: 552-61.
- Kelly, D. R. and T. Murray (1991a). "Twenty years of vocational training in the West of Scotland: the practice component." British Journal of General Practice 41(353): 492-5.
- Kelly, D. R. and T. S. Murray (1991b). "Twenty years of vocational training in the west of Scotland." BMJ(302): 28-30.
- Kelly, G. A. (1955). The Psychology of Personal Constructs. New York, Norton.
- Kember, D., D. Leung, et al. (2000). "Development of a Questionnaire to Measure the Level of Reflective Thinking." Assessment & Evaluation in Higher Education 25: 381-395.
- Ker, J. and D. Snadden (2001). Primary Care. A practical Guide for Medical Teachers. J. Dent and R. Harden. Edinburgh, Churchill Livingstone.
- Keynan, A., M. Friedman, et al. (1987). "Reliability of global rating scales in the assessment of clinical competence of medical students." Medical Education 21: 477-81.
- King, J. (2002). "360° appraisal." BMJ 324: 195.
- King, P. and K. Kitchener (1994). Developing Reflective Judgement: Understanding and promoting intellectual growth and critical thinking in adolescents and adults. San Francisco, Jossey-Bass.
- Koutsavlis, A. (2001). Disseminating Practice Guidelines to Physicians. Montreal, Direction de la Sante Publique.
- Kuyvenhoven, M., H. Pieters, et al. (1990). "The association between peer consultations and three aspects of clinical competence." Family Practice 7: 163-7.
- Litzelman, D., G. Stratos, et al. (1998). "Beneficial and harmful effects of augmented feedback on physicians' clinical-teaching performances." Academic Medicine. 73: 324-32.
- Lough, J. and T. Murray (1997). "Training for audit: lessons still to be learned." British Journal of General Practice. (418): 47: 290-2.
- Marinker, M. (1992). Assessment of Postgraduate Medical Education - Future Directions. General Practitioner Education, UK and Nordic Perspectives. M. Lawrence and P. Pritchard. London, Springer Verlag: 75-80.
- Marshall, M. and T. Wilson (2002). "Quality Supplement." British Journal of General Practice. 52: S1-S48.
- Martys, C. R. (1979). "Some trainees' views on vocational training." Update(18): 1079-1081.
- McGuire, C. (1983). Evaluation of student and practitioners competence. Handbook of Health Professions Education. C. McGuire, R. Foley, A. Gorr and R. Richards. San Francisco, Jossey Bass: 256-293.
- McKay, J., M. Lough, et al. (2002). "Registrar Audit and implementation of change: a pilot project." Education for Primary Care 13(3): 336-339.
- McKinley, R., R. Fraser, et al. (2001). "Model for directly assessing and improving clinical competence and performance in revalidation of clinicians." BMJ 322: 712-715.

Evaluating Process and Outcome in the Education of General Practitioners

- Munro, N., R. Hornung, et al. (1998). "What are the key attributes of a good general practice trainer?" Education for General Practice 9(2): 263-270.
- Munson, B. (1998). "Peers observing peers: The better way to observe teachers." Contemporary Education 69: 108-10.
- Neighbour, R. (1999). The Inner Consultation: How to Develop an Effective and Intuitive Consulting Style. London, Petroc Press.
- Newble, D., B. Jolly, et al. (1994). The Certification and Recertification of Doctors: Issues in the Assessment of Clinical Competence. Cambridge, Cambridge University Press.
- Norcini, J. J., D. Diserens, et al. (1990). "The Scoring and Reproducibility of an Essay Test of Clinical Judgement." Academic Medicine 65: 9.
- Parker, J. and L. Rubin (1966). "Process as content: Curriculum design and the application of knowledge." Chicago, Rand McNally.
- Parlett, M. (1991). The assessment of hearing-impaired children. The reflective turn: case studies in and on educational practice. D. Schön. New York, Teachers' College Press.
- Peabody, J., J. Luck, et al. (2000). "'Comparison of Vignettes, Standardized Patients, and Chart Abstraction: A Prospective Validation Study of 3 Methods for Measuring Quality.'" JAMA 283: 1715-1722.
- Peile, E. (2000). "New Ways of Working: New Ways of Learning: Action Research in the Weston Project." School of Education. Oxford, Oxford Brookes University.
- Peile, E. (2001). GP Registrar training - what has lasting value? The General Practice Jigsaw. S. Field. Oxford, Radcliffe.
- Peile, E., G. Easton, et al. (2000). "The year in a training practice: what has lasting value? Grounded theoretical categories and dimensions from a pilot study." Medical Teacher 23: 205-11.
- Peile, E. and N. Johnson (2002). "Reliability issues in the assessment of GP trainers.".
- Peile, E. and A. Slowther (2001). Research In Medical Education: Constructing An Ethical Framework (Abstract). AMEE Scientific Meeting, Berlin.
- Pitts, J., C. Coles, et al. (1998). "Exploring the introduction of a performance-based component in the certification and recertification of general practice trainers." Education for General Practice 9(3): 316-23.
- Plewis, I. (1997). Statistics in Education. London, Arnold.
- Pringle, M., T. Wilson, et al. (2002). "Measuring "goodness" in individuals and healthcare systems." BMJ 325: 704-7.
- Ram, P., R. Grol, et al. (1998). "Structure and Process: the relationship between practice management and actual clinical performance in general practice." Family Practice 15(4): 354-362.
- Ramsey, P., M. Wenrich, et al. (1993). "Use of peer ratings to evaluate physician performance." JAMA 269: 1655-60.
- RCGP (1972). The future general practitioner. London, British Medical Association.
- RCGP (1985). What Sort of Doctor? assessing quality of care in general practice. London, Royal College of General Practitioners.
- Regehr, G., H. MacRae, et al. (1998). "Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination." Academic Medicine. 73: 993-7.
- Rethans, J.-J. and L. Saebu (1977). "Do general practitioners act consistently in real practice when they meet the same patient twice? examination of intradoctor variation using standardised (simulated) patients." BMJ 314: 1170.

- Rigby, K. and J. Metzger (1992). "Assessing community attitudes toward medical practitioner authority." Australian Journal of Public Health 16: 403-12.
- Roland, M. (1999). "Quality and efficiency: enemies or partners?" Br J Gen Pract 49: 140-143.
- Ronalds, C., A. Douglas, et al. (1981). Fourth National Trainee Conference. Reports, recommendations and questionnaire. Exeter 1980. Occasional paper 18. London, Royal College of General Practitioners.
- Rose, R. (1983). Twelve Angry Men. Woodstock, Dramatic Publishing Company.
- Roth, R. (1989). "'Preparing the Reflective Practitioner: Transforming the Apprentice through the Dialectic'." Journal of Teacher Education 40: 31-35.
- Rothman, A. I., D. Blackmore, et al. (1997). "The use of global ratings in OSCE station scores." Advanced Health Science Education 1: 215-9.
- Schlegelmilch, B. and J. Carman (1993). "Satisfaction with primary health care and confidence in general practitioners: an exploratory comparison of US and UK students." Health Services Management Research 6: 89-98.
- Schlottmann, A. and N. H. Anderson (1995). "Belief revision in children: Serial judgment in social cognition and decision-making domains." Journal of Experimental Psychology, 21(5): 1349-1364.
- Schofield, T. P. C. and J. C. Hasler (1984). "Approval of trainers and training practices in the Oxford region: assessment." BMJ 288: 612-4.
- Schofield, T. P. C. and J. C. Hasler (1984). "Approval of trainers and training practices in the Oxford region: criteria." BMJ 288: 538-40.
- Schofield, T. P. C. and J. C. Hasler (1984). "Approval of trainers and training practices in the Oxford region: evaluation." BMJ 288: 688-9.
- Scholle, S., C. Weisman, et al. (2000). "Women's satisfaction with primary care: a new measurement effort from the PHS National Centers of Excellence in Women's Health." Womens Health Issues, 10: 1-9.
- Schön, D. (1983). The Reflective Practitioner: How Professionals Think in Action. New York, Basic Books.
- Schön, D. (1987). Educating the Reflective Practitioner: Towards a New Design for Teaching and Learning in the Professions. San Francisco, Jossey-Bass.
- Sen, A. (2002). "Health: perception versus observation." BMJ 324: 860-861.
- Shapiro, J. and Y. Talbott (1991). "Applying the concept of reflective practitioner to understanding and teaching family medicine." Family Medicine 23: 450-6.
- Short, N. L. (1987). "Vocational training for general practice in Dartford 1971-1982: an assessment." Medical Teacher(9): 193-199.
- Slotnick, H. (1999). How Doctors Learn. Annual Meeting, Sussex, UK.
- Smith, F., A. Singleton, et al. (1998). "General Practitioner's continuing education: a review of policies, strategies and effectiveness, and their implications for the future." British Journal of General Practice 48: 1689-1695.
- Smith, W. (2000). "Evidence for the effectiveness of techniques to change physician behaviour." Chest 118: 8S-17S.
- Snadden, D. and M. Thomas (1998). "Portfolio learning in general practice vocational training--does it work?." Medical Education 32: 401-6.
- Snijders, T. and R. Bosker, , (1999). Multilevel Analysis: an Introduction to Basic and Advanced Multilevel Modelling. London, Sage.
- Southgate, L. (1994). "Freedom and discipline: clinical practice and the assessment of clinical competence." British Journal of General Practice 44: 87-92.

- Southgate, L., J. Cox, et al. (2001b). "The assessment of poorly performing doctors: the development of the assessment programmes for the General Medical Council's Performance Procedures." Medical Education 35 (Suppl. 1): 2-8.
- Southgate, L. and D. Dauphinee (1998). "Maintaining standards in British and Canadian medicine: the developing role of the regulatory body." BMJ 316: 697-97.
- Spencer, J. and R. Jordan (1999). "Learner-centred approaches in medical education." BMJ 318: 1280-3.
- Strauss, A. and J. Corbin (1998). Basics of Qualitative Research - techniques and procedures for developing grounded theory. Thousand Oaks, California, Sage.
- Streiner, D. (1985). Global rating scales. Assessing clinical competence. N. G. Nufeld VR. New York, Springer: 119-141.
- Tabachnick, B. and L. Fidell (1996). Using Multivariate Statistics. New York, Harper Collins.
- Thornham, J. R. (1980). "A survey of ex-trainees." Journal of Royal College of General Practitioners(30): 725-728.
- Tracey, J. M. (1997). "The validity of general practitioners' self assessment of knowledge: cross sectional study." BMJ 315: 1426-8.
- UDACE, Unit for Continuing Adult Education (1989). Understanding Competence. Leicester, National Institute for Adult Continuing Education (NIACE).
- Van der Vleuten, C. (1996). "The assessment of professional competence: developments, research and practical implications." Advances in Health Sciences Education 1: 41-67.
- Van Leeuwen, Y. D., S. Mol, et al. (1995). "Change in knowledge of general practice during their professional career." Family Practice(12): 313-7.
- Violato, C., J. Lockyer, et al. (2003). "Multi source feedback: A method of assessing surgical practice." BMJ(326): 546-548.
- Violato, C., A. Marini, et al. (1997). "Using peers, consulting physicians, patients, co-workers and self to assess physicians." Acad Med 70: 57-63.
- Walford, G. (2001). Doing Qualitative Educational Research: a personal guide to the research process. London, Continuum.
- Wensing, M., H. Jung, et al. (1998). "A Systematic Review of the literature on patient priorities for General Practice care. Part 1: Description of the Research Domain." Social Science Medicine 47: 1573-1588.
- Wensing, M., T. van der Weijden, et al. (1998). "Implementing guidelines and interventions in General Practice: which interventions are effective?" British Journal of General Practice 48: 991-997.
- While, R. and M. Attwood (2000). Professional Development: a guide for General Practice. Oxford, Blackwell Science.
- Whitfield, M. J. (1966). "Training for general practice: result of a survey into the general practitioner trainee scheme." British Medical Journal 1: 663-667.
- Wolf, M., S. Putnam, et al. (1978). "The Medical Interview Satisfaction Scale: development of a scale to measure patient perceptions of physician behavior." Journal of Behavioral Medicine. 1: 391-401.
- Wood, L. and E. O'Donnell (2000). "Assessment of competence and performance at interview." BMJ 320: S2-7321.
- Wright, S. (1996). "Examining what residents look for in their role models." Academic Medicine 71: 290-292.
- Wright, S., D. Kern, et al. (1998). "Attributes of Excellent Attending-Physician Role Models." New England Journal of Medicine 339: 1986-1993.

Appendix 1

GLOSSARY: Acronyms, Abbreviations and Definitions

APD	Accredited Professional Development
BMA	British Medical Association
CME	Continuing Medical Education
CPD	Continuing Professional Development
CHI	Commission for Health Improvement
FBA	Fellowship by Assessment
GMC	General Medical Council
GMSC	General Medical Services Committee of the BMA (now =GPC)
GPC	General Practitioners Committee of the BMA (formerly = GMSC)
HPE	Higher Professional Education
JCPTGP	Joint Committee on Postgraduate Training for General Practice
MAP	Membership by Assessment
PCT	Primary Care Trust
PGMDE	Departments for Postgraduate Medical and Dental Education
NCAA	National Clinical Assessment Authority
QPA	Quality Practice Award
QTD	Quality Team Development
RCGP	The Royal College of General Practitioners
VTs	Vocational Training Schemes
Appraisal	A formative review of performance and competence, for the benefit of the appraisee, resulting in the setting of educational objectives
Assessment	May be formative for the doctor assessed, as well as summative. Is owned by the assessor, and may be intended for the benefit of the organisation as well as ensuring that public expectations of standards are met
Associate Adviser	Educationalist in Postgraduate Deanery reporting to Director of General Practice Education, and responsible for a 'patch', e.g. county of Course Organisers or with a specific remit in medical education
Colleagues	Include not only fellow doctors, but all who work together in the Primary Care Team, including nurses, therapists and receptionists
Competence	What a doctor is capable of doing (Grol, Mookink et al. 1989)
Course Organiser	Educationalist in Postgraduate Deanery reporting to Associate Adviser. Responsible alone, or in partnership with another Course Organiser, for a locality Vocational Training Scheme, including running the Day Release Programme
Criteria	A discrete, definable, and measurable phenomenon, relevant to the definition of quality, and so clearly defined that we can say whether it is present or not (Donabedian)

Evaluating Process and Outcome in the Education of General Practitioners

Guideline	A systematically developed statement to assist decisions for practitioner and patient about appropriate health care for specific clinical circumstances (Institute of Medicine)
Indicator	A measurable element of practice performance for which there is evidence or consensus that it can be used to assess the quality, and hence the change in the quality, of the care provided (Lawrence, Olsen and Equip)
Perception	An opinion, based on experience, and interpreted subjectively, consisting of more-or-less subconscious judgements about performance
Performance	What a doctor actually does in daily practice (Grol, Mokkink et al. 1989)
Performance Management	A process for monitoring and reviewing substandard performance with the objective of raising the standard, but with the explicit understanding that in the event that performance does not improve, executive action will be taken
Standard	The level of compliance with a criterion (Black)

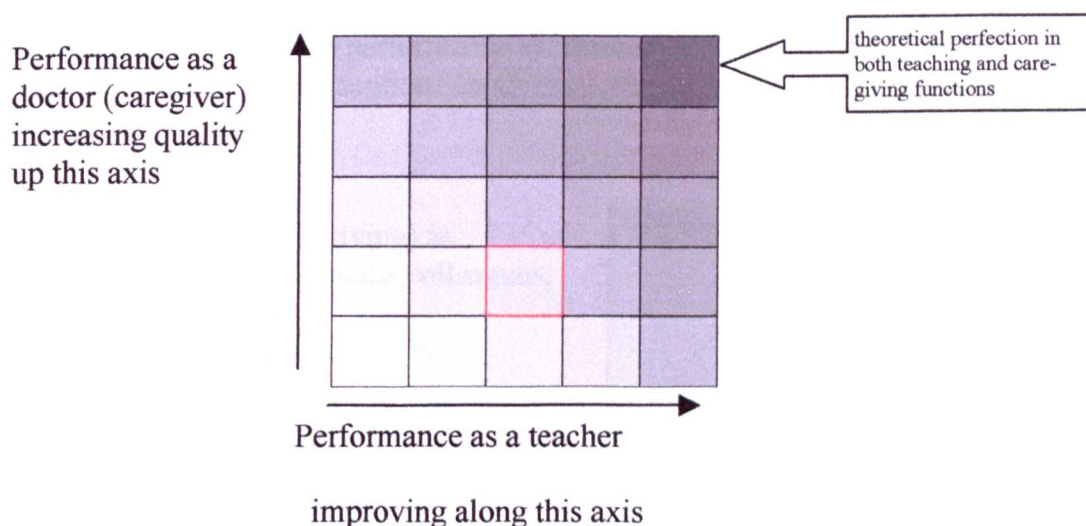
Appendix 2

Components of the conceptual model on perceptions

Appendix 2.1

Conceptual model A: The Trainer's performance; a notional reality matrix

Here is represented the Trainer's actual performance (if we could see reality this is what it would look like), in two dimensions: the function of the Trainer as a doctor caring for patients and the function of the Trainer as an educator training a Registrar.



The doctor who is performing at the level indicated in the square outlined in red is performing in the middle quintile as a teacher, but only on the second quintile as a doctor (caregiver).

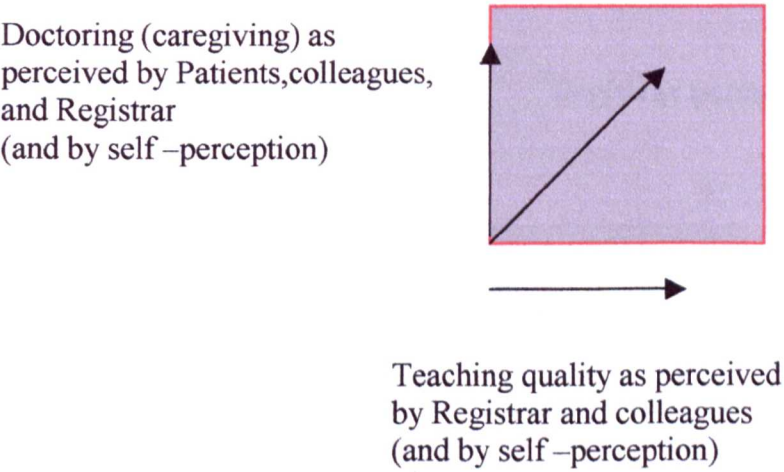
Appendix 2.2
Conceptual model B: Perceptions of the Trainer’s qualities in 2 dimensions

The Trainer’s performance influences the perception of those on whom he acts and those who are close enough to observe:

Doctoring (caregiving)	Teaching
<i>acts on</i> Patients	<i>acts on</i> Registrar
<i>observed by</i> Colleagues	<i>observed by</i> Colleagues
<i>observed by</i> Registrar	<i>(observed by</i> Patients)

NB although patients do observe some of the teaching, they are rarely asked to consider their perceptions of teaching

Consider then the doctor whose performance is represented in the square outlined in red on model 1.0. Relevant perceptions about his performance can be represented thus:



Appendix 2.3

Conceptual model C: Perceptions of the Trainer's performance; a notional reality matrix

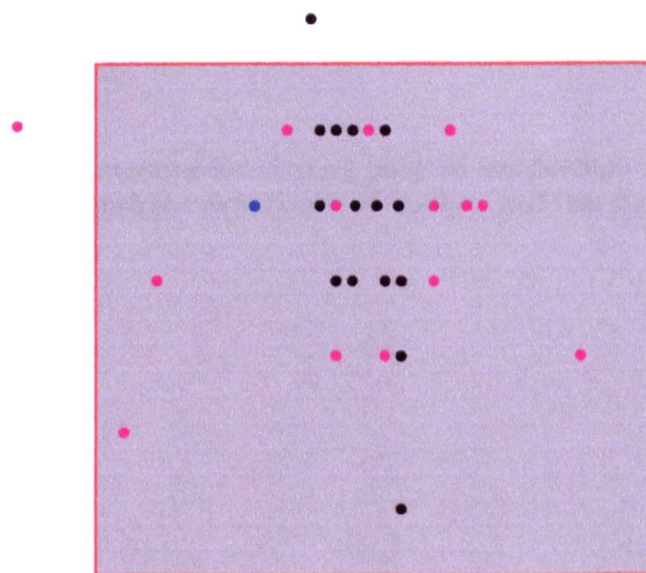
This model looks at 'real-time' experience or experiences of the Trainer's performance by other actors and observers.

If that experience is accurately reflected in their perception, then a plot of the immediate perceptions will reflect the Trainer's consistency in performance.

Consider, for example, 15 colleagues, 15 patients and 1 Registrar's experiences of the Trainer. In this illustration:

- Of the colleagues, 13 perceive the Trainer within the area of the aforementioned square, with some variation about how they rate his qualities along the axes of doctoring and teaching.
- The 15 Patients, have less perception of teaching qualities (so here this is represented as clustered around the mid-point of the x axis, on the principle that if they were asked, the patients would be likely to rate teaching as average, as they do not have extensive reference points of medical teaching). They vary up and down the y axis in how they rate the doctoring/caregiving, but, with one exception, they agree in their perceptions that Trainer quality lies within this box. There is only one Registrar represented in this model.

Patient perceptions = ● Colleague perceptions = ● Registrar perceptions = ●



NB Remember the statistical principle that the higher the number of observations, the greater the confidence we can have that the 'true' result is represented – that we are not looking at evidence of freak behaviour which is unrepresentative of actual performance. The evidence from 360 degree perceptions is that 'saturation' is achieved by 15 perceptions from any one group (Griffin, 2000).

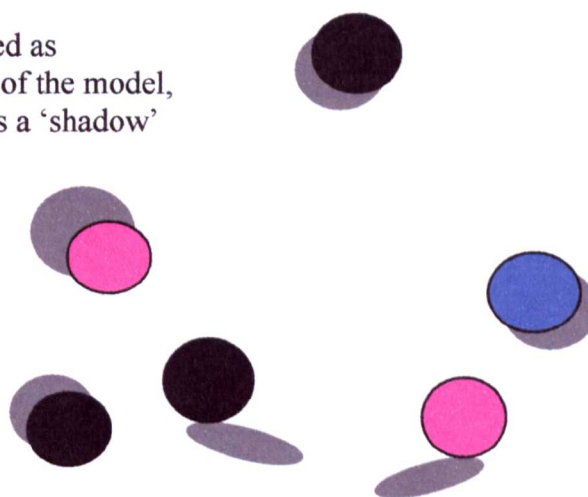
Appendix 2.4

Conceptual model D: Others' perceptions of the Trainer's performance; 'Reality' and the 'Personal Shadow'

This model is intended to show a 'shadow' effect around our perceptions of reality. (Sen, 2002) talks about the patient's self-**perception** of illness in a social context being affected by 'levels of education, availability of health facilities, and public information on illness and remedy'. We can likewise expect colleagues, patients and Registrars alike to interpret their experiences with the Trainer in the light of past experiences and personal constructs (Kelly, 1955).

This can be represented as a 'shadow' behind the perception. The 'Shadow' moves the perception away from our 'notional reality' in a direction that is determined by the individual characteristics of the perceiver.

Each individual perception, represented as a coloured dot in the previous section of the model, is here magnified to show how there is a 'shadow' for each one.



Effectively, our 'shadows' determine the starting point of our perceptions. Note that shadows can enlarge or diminish the experience of 'reality' and that the shape can get distorted.

Appendix 2.5

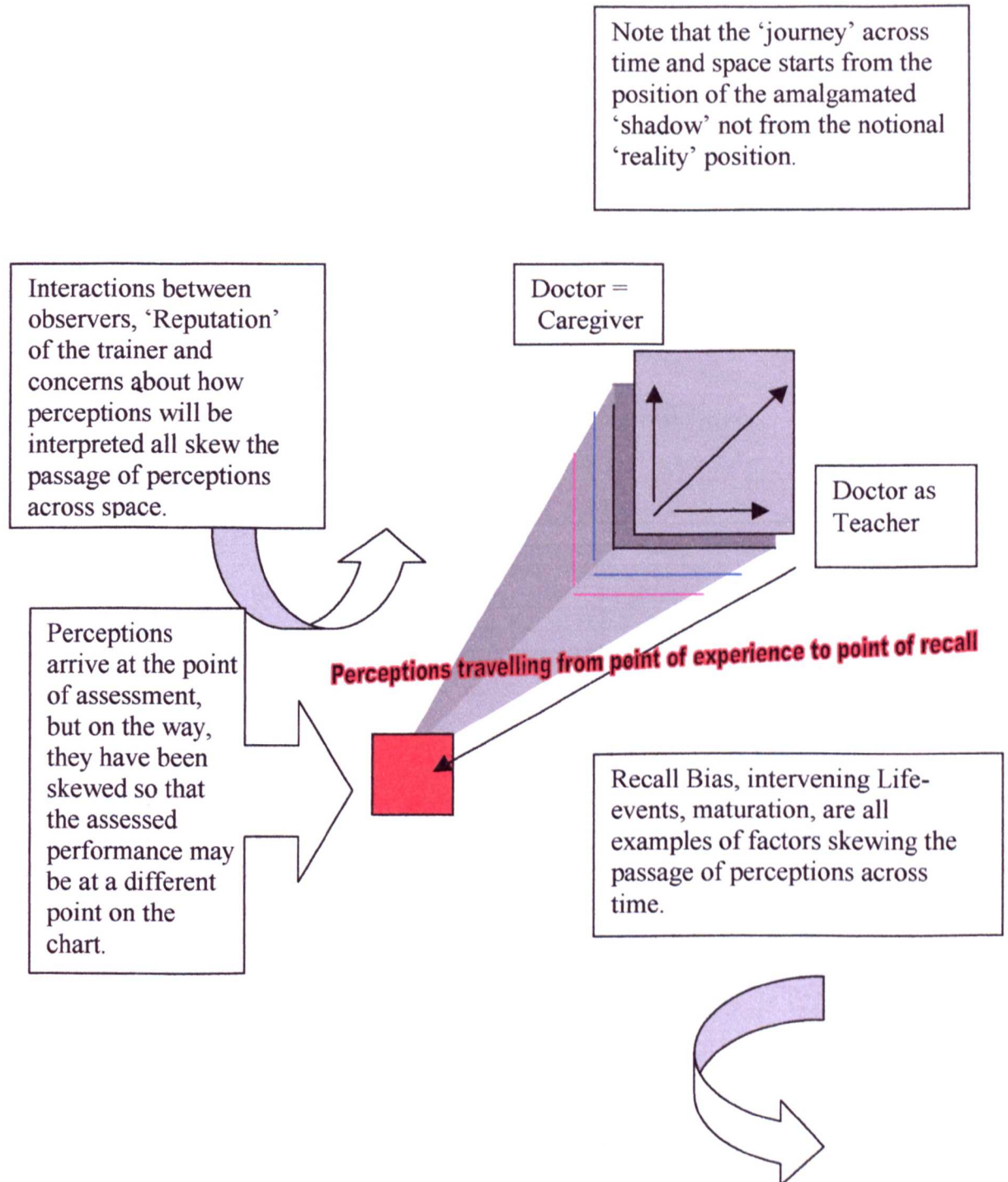
Conceptual model E: Perceptions of the Trainer's performance; travelling, not static

This element of the model introduces the concept that perceptions have to 'travel' before they reach those who are interested in assessing the Trainer's performance. In an ideal world, there would be no 'space' between the assessors and the Trainer they are assessing: they would assess his performance in 'real time' and 'on-the-spot'. What actually happens is that judgements are made in regional and national offices on performance over a period of years in a remote locality setting. Hence the notion that any **Perceptions**, on which these **Assessments** are based, have had to 'travel' across time and space.

In the course of this journey these assessments will be affected by numerous factors, some of which will affect only the perceptions of an individual, other factors may affect much of the group.

Evaluating Process and Outcome in the Education of General Practitioners

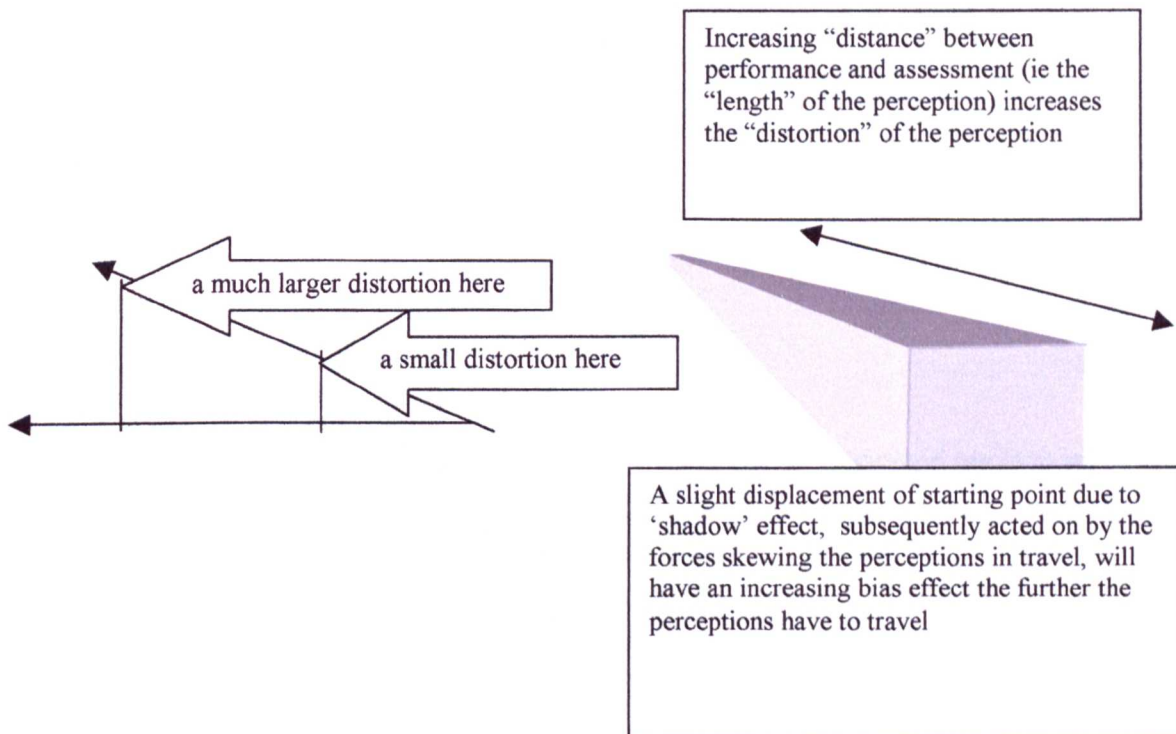
We can represent the dynamic group perceptions as a rectilinear structure.



Appendix 2.6

Conceptual model D: Others' perceptions of the Trainer's performance; the effect of distance

Again, distance can be conceived in a temporal or spatial dimension.

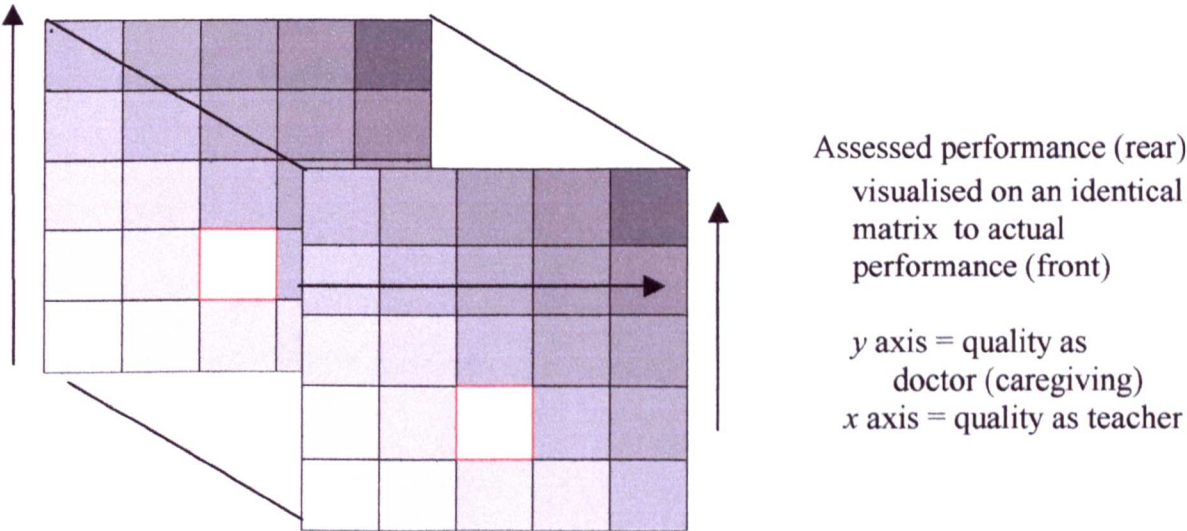


Appendix 2.7
Conceptual model G: Assessment of the Trainer's performance

Those who assess the Trainer's performance do so at a distance, and to a greater or lesser extent, their assessment is based on the evidence of perceptions of the Trainer's patients, colleagues, and Registrar.

We can conceive that the assessors work from an image of the Trainer's performance, which has an identical matriceal framework to that of actual performance. The assessors interpret perceptive evidence in the same dimensions of quality as a doctor (caregiver) and quality as a teacher.

This illustration shows the assessor's "correctly" placing the assessed performance of the Trainer in the same position as actual performance. This would constitute a valid assessment and were this usually the case, then the assessment would have a high degree of reliability.

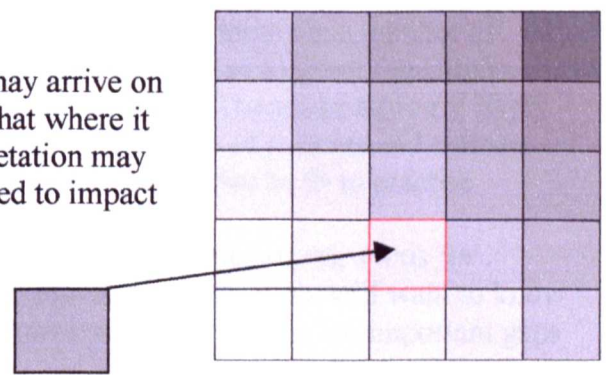


Appendix 2.8
Conceptual model H: Assessment of the Trainer's performance- factors
distorting perception-based assessment

Just as the **perceivers** have 'shadows', so do the **assessors**. They have to interpret evidence of perceptions and in this exercise of interpretation there will be an affect of all past experiences, preconceptions and personal constructs.

Thus, the cuboid column of perceptions may arrive on the assessors' grid at a different point to that where it originated, and also the assessors' interpretation may further shift the point at which it is reported to impact

Where will the shadow' lie?



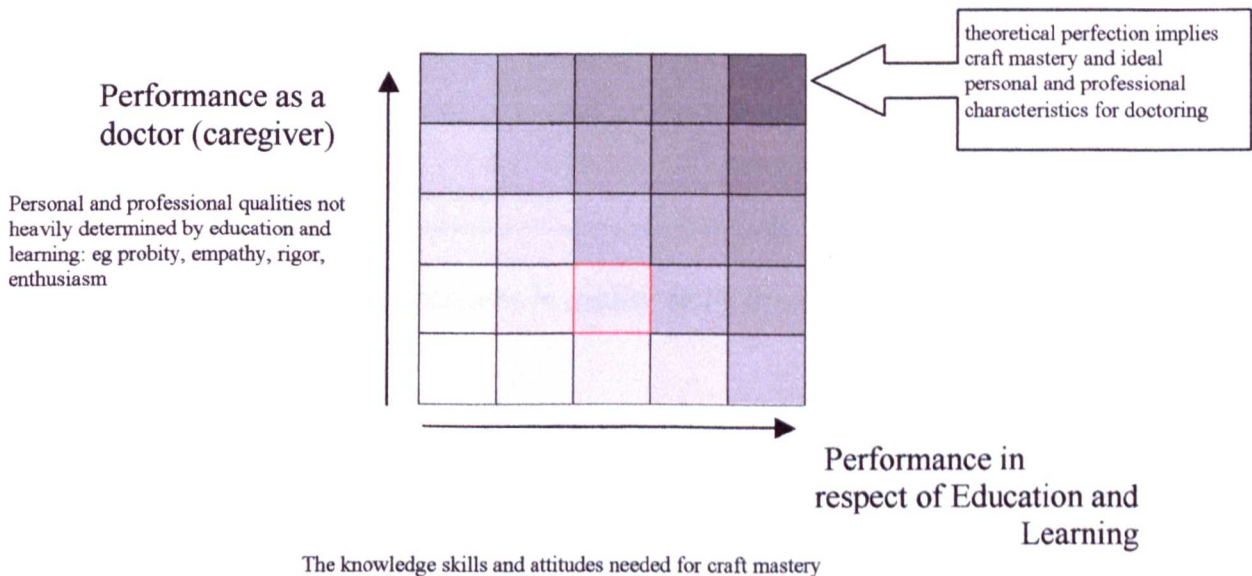
Appendix 2.9

Conceptual model I: Performance, Perceptions and Assessment – variations on a theme – assessing the trained Registrar

In the model, so far, we have concentrated on the assessment of the Trainer, whether it be by organisations such as PGMDE (primarily interested in his qualities as an educator) or by ones such as the GMC (primarily interested in his qualities as a doctor).

A similar model can be used to look at the Registrar. If there are concerns about the trained Registrar, working as an independent practitioner, there are a number of bodies who may be required to make an assessment of her as a doctor (eg Clinical Governance at the PCT, CHI, NCAA, or even the GMC). These assessors are likely similarly to base much of their assessment on the evidence of patients and colleagues, and probably also the Trainer who has ‘signed off’ the doctor as fit to practise.

The dimensions remain similar: y axis for the caregiving dimension, x axis for education, but there are some differences in the latter, as assessors will want to know if the doctor has adequately mastered the craft learning or if there are important gaps in her medical education.



Appendix 2.10

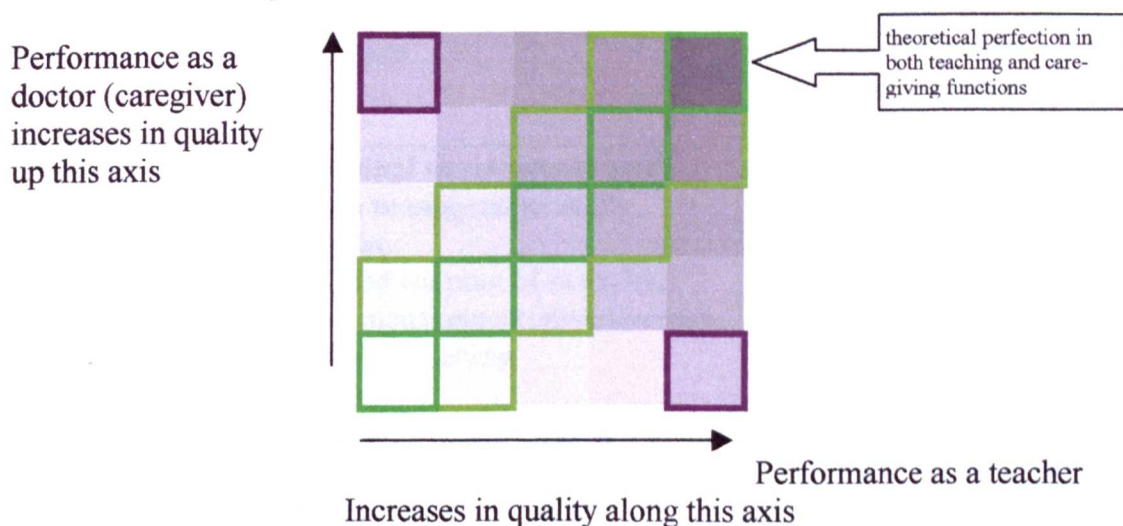
Conceptual model J: Performance, Perceptions and Assessment – interactions between the ‘Education’ and the ‘Caregiving’ dimension

The matrix grid introduced in Conceptual Model 1.0 might be taken to imply that Education and caregiving are independent qualities, but of course they are not. They interact in a number of different ways: perhaps most importantly in their dependency on common skills and attitudes. For example:

- Communication skills
- Empathy
- Interest in people

All of these are determinants of the qualities of the teacher and the caregiver, as is brought out when new Trainers realise that they can use their consulting skills as teaching skills, substituting learner-centeredness for patient-centeredness.

This means that we might expect an association between both qualities, which we can represent on the matrix grid thus:



The bright green lines outline the cells where we would expect to find the Trainer's performance if both qualities are directly related. The darker green lines outline the adjacent cells, denoting performance wherein the two qualities are related to a lesser extent.

By this reasoning, outliers in the purple edged cells ought to give us pause for thought: how come someone manages to perform brilliantly as a teacher but abysmally as a doctor, or vica-versa?

The same considerations apply to trained Registrars (Conceptual Model 6.0) for different reasons. Doctors with high levels of probity, rigor, enthusiasm, usually apply these qualities to their learning. A finding of total disassociation between the two dimensions should likewise give us pause for thought.

**PAGE
MISSING
IN
ORIGINAL**

Appendix 3

Categories and Dimensions of Educational Behaviours by GP Trainers that have lasting value for trained doctors

		Preferred Behaviour	Less helpful Behaviour
1	Training or Education	Problem-Based Approach Teaching based on approaches to problems which are not limited to present-day contexts	Emphasis on Managing Disease Teaching focused on current policies for disease management
2	Style Spectrum	Wide variety of styles Learner exposed to different consulting styles and role-models in tutorials	Narrow range of styles Teaching dominated by personal style and behaviour of Trainer
3	Space for Reflection	Encouraging reflective practitioner Safe environment to learn from mistakes	Protocol driven behaviour Black and white approach adopted where learner is expected to adhere to guidelines and elements of blame culture likely.
4	Modelling Personal Development and Team Skills	Personal development and team management skills taught Guided learning of skills like time management, assertiveness, boundary-setting	No emphasis on team behaviours Little attempt is made to help learner understand the importance of team-working and the areas of personal development that are involved
5	Learning Cycles	Learning cycles completed A culture exists in the practice where reflection, audit, assessment all promote change and re-evaluation	Haphazard change Culture is reactive to external pressures, and little evidence of information about the practice inspiring meaningful change
6	Family practice in context	Contextualised Learning Trainer introduces the broader dimensions of family and health expectations	Emphasis on presenting problem Focus remains on sorting and shifting
7	Control and Direction	Learner-centred approach Trainer encourages progression towards self-directed learning	Trainer-centred approach Trainer fails to match style to learner's needs. Eg Rigid structure with fixed diet or total lack of direction.
8	Feedback	Sensitive feedback Both positive and negative feedback delivered where appropriate, stimulating confidence in the learner, and encouraging change.	Inappropriate criticism Feedback either inadequate or misplaced or poorly delivered, often not timely or specific enough to be useful to learner.

Appendix 4

Trainers' information about assessment visits

Dear Trainer

I thought I would let you know about an educational research activity, which will be going on during your forthcoming accreditation/re-accreditation visit. This activity, which is essentially a change in the way some of the information is recorded, will not directly affect the outcome of your accreditation/re-accreditation, which is still based on the criteria with which you are familiar.

However, the visitors are experimenting with a new way of recording information about training behaviours based on some research conducted by Dr Ed Peile and others. Some people have commented that, up to now, visits have had more of an emphasis on the structure of the training and training practice rather than the process of the education that's delivered and we are looking at ways of putting more emphasis on the process of training.

You can expect little change in the organisation of the visit. The only change you may notice is that, in some sections of the visit: the watching of the video tutorial, the Trainer interview, the Registrar interview, and perhaps when looking at some of the records of training, one or more of the visiting team may be jotting some notes on some structured forms. At the end of the training visit, possibly after the feedback to the practice, the visiting team will compare their notes and the Team-leader will then be in a position to feedback to you something of what they've gleaned about training behaviours in the hope that this may give you some useful material to reflect on. The interest in training behaviours has stemmed from earlier work in which former GP Registrars have discussed what has had lasting value for them in their subsequent work as trained GPs.

If you have any feedback about any perceived effects of this research activity at the time of your visit please, do not hesitate to contact me or contact Ed Peile directly (phone/fax 01296 631727, email: ed.peile@dphpc.ox.ac.uk) as we are very interested in the positive and negative aspects of increasing the emphasis on training behaviours at re-accreditation visits.

Yours sincerely
Director of Postgraduate General Practice Education.

Appendix 5

Teamleaders' plans for Phase IIc

Assessing training behaviours on reaccreditation visits

Introduction

The purpose of this research work is to move us further forward into the area of looking at Trainer behaviours. Our pilot work has shown that Trainers really welcome meaningful exploration of their teaching behaviours and feedback on this and in due course, if properly researched, it may be appropriate for some educational behaviours to influence the reaccreditation process.

The methods that you will be using are all methods that are currently in use on reaccreditation visits with the exception of standard scenarios.

Collecting the evidence

It is important to emphasize that there is virtually no change to the existing visit format. Just ask the questions as you have been doing up to now. All that will be different is the way that you will record some of the evidence.

Recording the evidence

It is suggested that for each section of the visit one team member is tasked with recording the evidence on training behaviours onto the worksheets. Other team members can continue as normal.

In order to do this you will need to make sure that before the visit you are familiar with the categories of training behaviours and their dimensions spanning what we call 'the preferred behaviour' to what we call 'the less helpful behaviour'. If you are familiar with the eight categories it will become easy to record evidence on the work sheets.

What we suggest you do is just jot down a little *aide memoir* of any point which seems to be evidence. Make a quick decision as to where you see this evidence lying across the dimension - is it neutral or is it strong or very strong evidence of a preferred behaviour, or does it appear that this is a less helpful behaviour and should go under the weak or very weak column? Don't agonise about this, just put it in where it seems right - you can change it later at team discussion.

Evaluating Process and Outcome in the Education of General Practitioners

Team discussion

You will need to allow a little bit of time at this point in the visit for team discussion. Those who have been recording the evidence need to collate it together and the team needs to reach a consensus as to whether the evidence, overall, amounts to very strong, strong, intermediate, weak or very weak, for the particular category. We don't want to give you guidance on this, just use your own judgement on weighting the evidence that you've recorded.

Feedback

It is very important that you model sensitive feedback on discussing training behaviours with the Trainer. You should negotiate whether you do this as a whole practice or with the Trainer individually but we are particularly interested, for the research exercise, in whether the Trainer found this feedback on training behaviours useful. We hope it will open up more discussion into a highly relevant area of training.

Data collection

When you have completed the four work sheets please return them to Simon Plint at PGMDE who will collate the information. Team-leaders are particularly asked to make comments on the team's experience of working with the new method and the helpfulness or otherwise to the Trainer and training practice.

Evaluating Process and Outcome in the Education of General Practitioners

1	Training or Education Ref: 1,2,3,4, (see below)	Problem-Based Approach Teaching based on approaches to problems which are not limited to present-day contexts (preferred behaviour)	Emphasis on Managing Disease Teaching focused on current policies for disease management (less helpful behaviour)
2	Style Spectrum Ref: 1,3,4 (see below)	Wide variety of styles Learner exposed to different consulting styles and role-models in tutorials (preferred)	Narrow range of styles Teaching dominated by personal style and behaviour of Trainer (less helpful)
3	Space for Reflection Ref: 2,3,4, (see below)	Encouraging reflective practitioner Safe environment to learn from mistakes (preferred)	Protocol driven behaviour Black and white approach adopted where learner is expected to adhere to guidelines and elements of blame culture likely (less helpful)
4	Modelling Personal Development and Team Skills Ref: 1,2,3,4 (see below)	Personal development and team management skills taught Guided learning of skills like time management, assertiveness, boundary-setting (preferred)	No emphasis on team behaviours Little attempt is made to help learner understand the importance of team-working and the areas of personal development that are involved (less helpful)
5	Learning Cycles Ref: 1,3,4 (see below)	Learning cycles completed A culture exists in the practice where reflection, audit, assessment all promote change and re-evaluation (preferred)	Haphazard change Culture is reactive to external pressures, and little evidence of information about the practice inspiring meaningful change (less helpful)
6	Family practice in context Ref: 2,3,4 (see below)	Contextualised Learning Trainer introduces the broader dimensions of family and health expectations (preferred)	Emphasis on presenting problem Focus remains on sorting and shifting (less helpful)
7	Control & Direction Ref: 1,2,3,4 (see below)	Learner centred approach Trainer encourages progression towards self-directed learning (preferred)	Trainer centred approach Trainer fails to match style to learner's needs. Rigid structure, fixed diet (less helpful)
8	Feedback Ref: 1,2,3,4 (see below)	Sensitive feedback Both positive and negative feedback delivered where appropriate, stimulating confidence in the learner, and encouraging change (preferred)	Inappropriate criticism Feedback either inadequate or misplaced or poorly delivered, often not timely or specific enough to be useful to learner (less helpful)

Methods for obtaining evidence

1. Log and Programme
2. Video of tutorial
3. Registrar interview
4. Trainer Interview

Evaluating Process and Outcome in the Education of General Practitioners

Methods for obtaining evidence

1. Log and Programme

- emphasis on managing disease or problem-based approach?
- exposure to other doctors and other team members throughout the year?
- opportunities used for learning about self and teams?
- Time set aside for teaching audit and team skills?
- Programmed personal development teaching?
- evidence of movement towards self direction?
- Feedback recorded and tracked?

Audit

- evidence of understanding importance of completing audit cycles
- use of learner's project and previous learners' projects

Practice protocols & guidelines

- reflective practice encouraged

2. Video of tutorial

- emphasis on managing disease or problem-based approach?
- reflective practice encouraged?
- opportunities used for learning about self and teams?
- evidence of family contextualisation?
- ad-hoc evidence of Trainer adapting to learner stage?
- sensitive feedback demonstrated?

3. Registrar interview

- emphasis on managing disease or problem-based approach?
- exposure to other doctors and other team members throughout the year?
- reflective practice encouraged?
- opportunities used for learning about self and teams?
- learner aware of completed audit cycles in practice, and resultant change management?
- learner understands educational needs assessment?
- evidence of Trainer encouraging learner direction within Trainer control?
- comfortable about feedback and able to give examples of pos/neg?

4. Trainer Interview

- emphasis on managing disease or problem-based approach?
- reflective practice encouraged?
- opportunities used for learning about self and teams?
- Needs assessment a part of care process?
- Needs assessment a part of educational process?
- evidence of Trainer encouraging learner direction within Trainer control?
- comfortable about feedback and able to give examples of pos/neg?

Training or Education

1	Problem-Based Approach Teaching based on approaches to problems which are not limited to present-day contexts (preferred behaviour)		Emphasis on Managing Disease Teaching focused on current policies for disease management (less helpful behaviour)
Metho ds Log Video Registr ar Trainer	Preferred approach	Intermediate/Neu tral	Less helpful behaviours

Style Spectrum

2	Wide variety of styles Learner exposed to different consulting styles and role-models in tutorials (preferred behaviour)		Narrow range of styles Teaching dominated by personal style and behaviour of Trainer (less helpful behaviour)
Metho ds Log Registr ar Trainer	Preferred approach	Intermediate/Neutr al	Less helpful behaviours

Evaluating Process and Outcome in the Education of General Practitioners

Space for Reflection

3	Encouraging reflective practitioner Safe environment to learn from mistakes (preferred behaviour)		Protocol driven behaviour Black and white approach adopted where learner is expected to adhere to guidelines and elements of blame culture likely (less helpful behaviour)
Methods Log Video Registrar Trainer Protocols Other	Preferred approach	Intermediate/Neutral	Less helpful behaviours

Modelling Personal Development and Team Skills

4	Personal development and team management skills taught Guided learning of skills like time management, assertiveness, boundary-setting (preferred behaviour)		No emphasis on team behaviours Little attempt is made to help learner understand the importance of team-working and the areas of personal development that are involved (less helpful behaviour)
Methods Log Video Trainer Registrar	Preferred approach	Intermediate/Neutral	Less helpful behaviours

Learning Cycles

5	Learning cycles completed A culture exists in the practice where reflection, audit, assessment all promote change and re-evaluation (preferred behaviour)		Haphazard change Culture is reactive to external pressures, and little evidence of information about the practice inspiring meaningful change (less helpful behaviour)
Methods Log Registrar Trainer Audit	Preferred approach	Intermediate/Neutral	Less helpful behaviour

Family practice in context

6	Contextualised Learning Trainer introduces the broader dimensions of family and health expectations (preferred behaviour)		Emphasis on presenting problem Focus remains on sorting and shifting (less helpful behaviour)
Methods Std scenario Registrar Trainer Log Video	Preferred approach	Intermediate/Neutral	Less helpful behaviours

Evaluating Process and Outcome in the Education of General Practitioners

Control and Direction

7	Learner centred approach Trainer encourages progression towards self-directed learning (preferred behaviour)	Trainer centred approach Trainer fails to match style to learner's needs. Rigid structure -fixed diet (less helpful behaviour)
Methods Registrar Video Trainer Log	Preferred approach	Less helpful behaviours

Intermediate/Neutral

8		Inappropriate criticism Feedback either inadequate or misplaced or poorly delivered, often not timely or specific enough to be useful to learner (less helpful behaviour)
Methods Registrar Log Trainer Video		Less helpful behaviours
	Intermediate/Neutral	

Appendix 6

Team-leaders' study day revising framework

**Team-leaders meeting on Use of Framework to look at Educational Behaviours
on re-accreditation visits. 8th November, 2001**

Executive Summary

1. Team-leaders support the use of the framework and wish to see it become a more central part of re-accreditation visits.
2. At present it is unwieldy and very difficult to incorporate collecting evidence for the framework into re-accreditation visits when there is so much work to be done running the visit to the existing format.
3. Radical change to visit format is proposed with the emphasis of the visits being on the doctor (and other educators) as an educator.
4. The other components of the visit agents (Doctor as Trainer: Practice as Service Deliverer: Practice as Training Environment) will be largely assessed on evidence provided before the visit by the Trainer and Practice. Only a small amount of visit time should be taken up in verifying these items.
5. Trainers should be invited to train as research associates, expert in collecting evidence for the new framework. Their knowledge would then be cascaded through the training community.
6. The new framework should be taught at new Trainers and experienced Trainers courses.
7. Research should be urgently commissioned to validate the decision making process using the new frameworks.
8. In the interim, teams will continue to operate the framework.
9. We will collect data on the best way of collecting evidence for the behaviours.

**Report of Team-leaders Meeting
8th November
Milton Keynes**

Implications for revision of Phase II of Registrar Project

John Toby, Chair of Joint Committee on post-graduate GP training nationally, opened the meeting by looking at the ways in which criteria for re-accreditation for training practises are being revised. This is dependent on the anticipated revision of Good Medical Practice for GPs which is currently being rewritten. This raised the thought should we realign our Insight 360 questions along GMP lines??

The main principle of re-accreditation provision is that there will be emphasis on the applicant providing evidence and the visit verifying the material to be assessed at re-accreditation falls into four categories:

- The doctor as a doctor (GMC compliant)
- Doctor as Trainer (Peile framework)
- Practice as Service Deliverer (Clinical Governance / Quality Practice Assessment)
- Practice as training environment

Team-leaders Discussion about using categories of training behaviours ('Peile framework')

Simon Plint started by asking Team-leaders to talk about their experience.

SP: Feels in is really promising, very enthusiastic, struggles on the day to use it, best experienced when tasked one team member to exclusively collect evidence. This only possible on visiting single Trainer practices. Finds the small grid of 8 categories on one page works better.

SS: Agrees with the practicality of using 2 mental frameworks is difficult. The framework was not done that well. Should we video the Trainer interview and record the evidence later. It takes time to explain to Trainers but worth it.

NT: The problems are the newness of being a Team-leader and the newness of Trainers to being team members. Team has many new Trainers and they cannot cope with learning two frameworks.

JM: Finds new framework very useful. Allocates particular person for particular task on visits. Team members not prepared to stay behind for feedback at end of long day. He stays to give feedback personally, finds Trainers really appreciated.

AC: 'Total and abject failure' to use criteria. Completely new concept. Difficult for teams to take on board. Therefore feedback been non-constructive. However remains keen in theory.

JT: Very mixed experience. Too much to do on the day particularly when multiple Trainers. Seen EP use framework effectively.

Evaluating Process and Outcome in the Education of General Practitioners

There followed general discussion and there was agreement:

1. The framework is promising.
2. Team-leaders do feel that it is an advantage to have a framework to look at educational behaviours, but they find that there is too much to do on the day.
3. The preferred option is rather than abandon the framework to radically change the visiting structure, so that the visit concentrates far less on doctor as doctor, practice as service deliverer, and practice as training environment – only setting aside an hour or so to verify previously reported information in these areas, and concentrates instead on looking at the doctor as Trainer using the new criteria.

SP had an idea that was welcome by the rest of the group.

The suggestion is that EP should advertise for Team-leaders wishing to train as research associates. They should be properly trained in the use of the new techniques on an accredited training day using videos, etc. and this module could also be used in the new Trainers course. One of these research-trained visitors could then be allocated to every visit and they could train other visitors and Team-leaders in the use of the categories.

At this point EP gave some information on the visit so far. Data quality from the research view point had been poor reflecting the difficulty that teams and Team-leaders had in handling the forms in visit days. However the few comments that had been received as feedback from Team-leaders had been more positive than negative. (see Attachment 1)

Team-leaders pointed out that the response from Trainers had also been very largely positive to the feedback they had received on their own educational behaviours.

The information on the different categories was interesting. (see Attachment 2)

The points to note are:

- Most Trainers are using preferred behaviours in the reflection category.
- The category where there is most scope for improvement is in use of 'Style' spectrum followed by 'Learning Cycles' and 'Personal Development' and 'Team Skills'. Categories where Trainers found it difficult to judge between 'Preferred behaviours' and 'Less Preferred Behaviours' (mixed categories) were 'Learning Cycles', 'Style' spectrum, 'Personal Development' and 'Team Skills' and 'Family Practice in Context'.

Validating the process:

It was widely acknowledged that although the categories and dimensions are well researched and well validated, the process of collecting the information and making judgements needs to be validated. It was at this point that the research protocol designed by Simon Street and Ed Peile was mentioned and it was felt that it was likely to be fruitful. (see Attachment 3)

The better validated process was seen to be important for the future in terms of defending any decisions that are made as a result of these visits.

Summative or Formative

There was a lively discussion about whether the use of a framework for looking at educational behaviours should be Summative or formative. Ed said that he felt that it should be formative for some time to come, and that he saw the process as perhaps formative assessments on a first and second visit, benching marking the progress that a Trainer had made in any one category and then when several Trainers had had repeat assessments, looking at whether we could make a Summative judgement here. John Toby disagreed. He thought there was potential to make this Summative more quickly and indeed a need to do so. He instanced examples of the progress made in different areas where Summative decisions were being reached which would previously have been thought to be impossible. "We should be Summative about what matters and this matters".

Conclusions and Decisions

1. The group wishes to retain the framework to operate it more effectively to validate it and to make it a major component of revisions of training visits.
2. We need to include more evidence particularly from the course organiser and particularly from the past Registrars' reports that will line up with the new framework.

Action

1. SP volunteered to help on design
2. With Director's approval, EP will advertise for Trainers to train as research associates to become skilled assessors in visits.
3. With Director's and Course Leader's approval, training on the new categories will be introduced into a new Trainer's course and higher Trainer's course. **Action: EP to liaise and design on this.**
4. The working group on the criteria for re-accreditation will incorporate the framework into the new criteria. **Action: JT to lead on this.**
5. The research proposed by Simon Street and Ed Peile should receive active support. **Action: SS/EP**
6. In the meantime, the visits will continue on the present format with data collection being as good as we can make it. It was felt that it would be retrograde to stop incorporation of the framework into visits now when there is a momentum behind it. **Action EP to encourage better data collection via Barbara Gow**
7. Simon Plint will lead work on collecting useful approaches and questions to access the information we need on visits. **Action: SP**

Attachment 1 to Appendix 6

1. How easy or difficult was it to incorporate this assessment into the routine visit?

I delegated the task to my colleagues. – a reflection of my own lack of preparation. I now feel we should have spent time to internalise these criteria before the interview.

Difficult – but first time.

Better second time around- definitely possible. Unfortunately neither of Trainers were familiar with categories so had to spend 20min introducing it at the beginning. They were not happy to record it so I did it.

Quite easy when familiar with the categories.

2. Did the emphasis on training behaviours have any beneficial or detrimental effects on the balance of emphasis of your visit?

No.

Helpful because there were a lot of issues about training.

Beneficial – More emphasis on the training experiences.

Helpful to concentrate our minds on teaching.

3. Do you think team members found categories and dimensions a useful framework once they had started to use them?

I don't think we really remembered them.

Yes.

Yes - they were very interested.

Yes useful for the visit and useful for reflection on own teaching.

4. Do you think the visit Trainer derived benefit from the structured feedback?

No.

More able to give specific example.

Yes, he was very positive.

Yes.

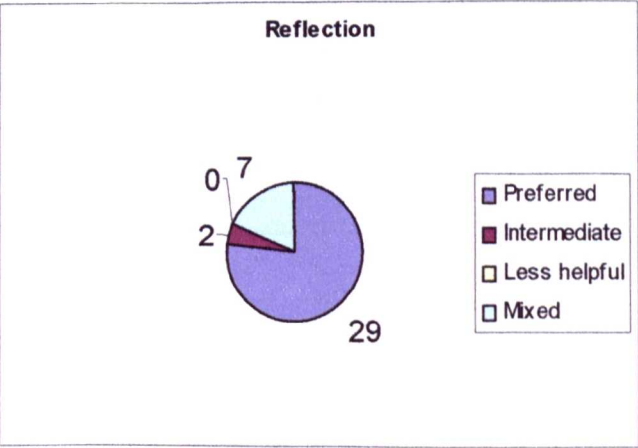
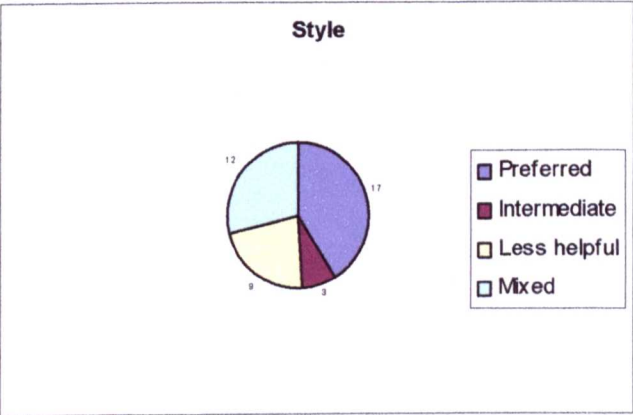
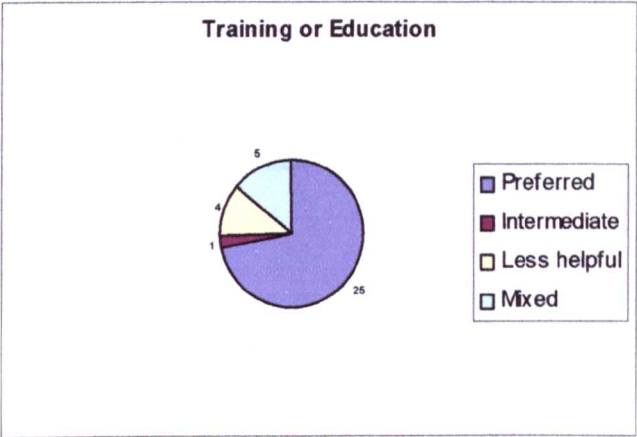
5. Any other comments?

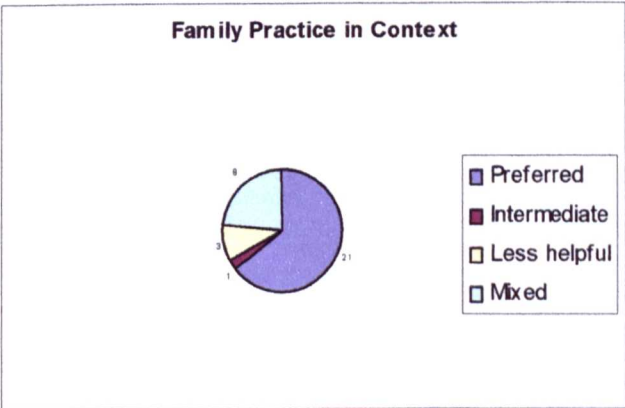
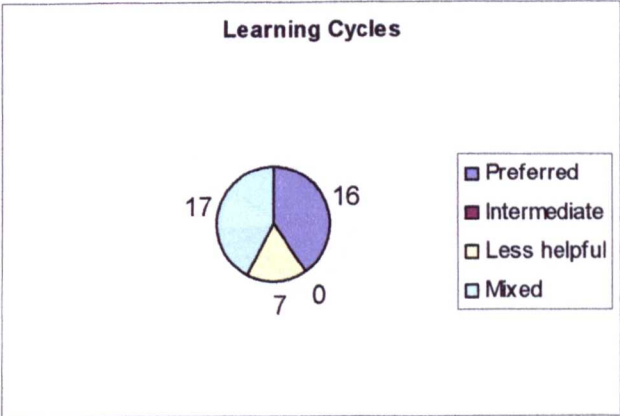
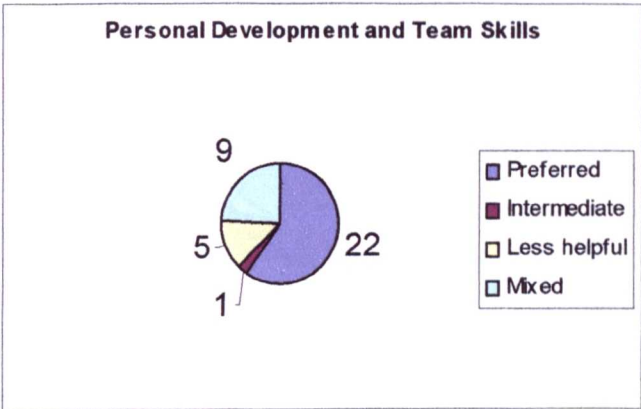
I've learned that this is not an experience which can be picked up on the day of the visit.

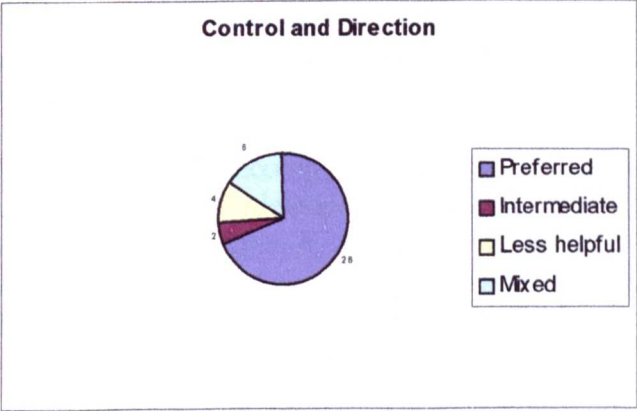
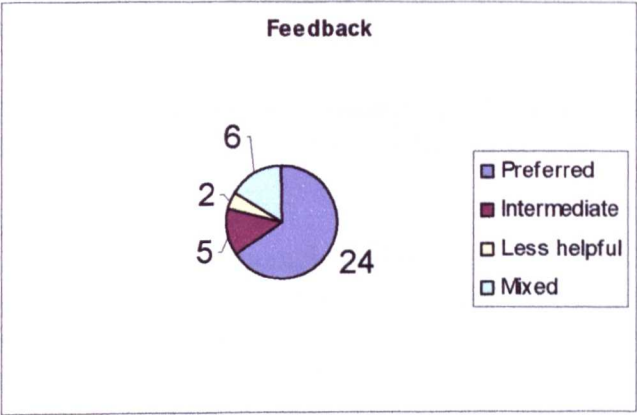
An extremely useful exercise. We need to watch a tendency to oversimplify the categories.

We are getting use to it now.

Attachment 2 to Appendix 6







Attachment 3 to Appendix 6

How do Peers Assess Training Behaviours on Re-accreditation Visits to GP Training Practices? A qualitative study

Summary

Until recently the educational behaviours used by Trainers of GP Registrars received scant attention at re-accreditation visits. A framework to assess 8 training behaviours across a dimension which spans 'Preferred Behaviours' to 'Less Helpful Behaviours' is now being trialled on re-accreditation visits throughout Oxford Deanery.

We aim to conduct a qualitative study of these assessments, to clarify how the judgements are made, and how Trainers who are assessed perceive the validity of the process and the formative value of the assessment of their training behaviours.

A researcher will accompany 6 different peer review teams on 6 visits to different Trainers, and will videotape the assessors' discussion on training behaviours, and the feedback to the Trainer in question. The researcher will subsequently interview team members and the Trainer individually, using Interpersonal Process Recall (IPR) (Kagan, 1991) to analyse the steps taken to arrive at decisions. Recordings of these interviews will be transcribed, and this material, together with transcripts of the videotapes, will be subjected to a grounded theoretical interpretation.

One reason why the actual educational process of training has so far eluded assessment, despite being the crux of a GP Trainer's work, is that we lack confidence in our ability to assess educational qualities and attitudes, underpinning behaviours. This research, by shedding light on the process by which peers make judgements, and on the confidence placed in the assessments, could enable training assessments to move forward into more relevant areas and to become a useful tool in the continuing development of GP Trainers.

Appendix 7

Attributes of an assessment process

Five required attributes of an assessment process

Reliability is a measure of the variation in scores due to differences in performance between subjects and also the correlation of assessors rating the same performance. It is generally accepted that the reliability of a regulatory assessment must be at least 0.8

Validity is the degree to which an assessment is a measure of what should be measured. Although face validity of an assessment (the extent to which an assessment measures what it purports to measure) is often discussed, this should be augmented by discussion of whether what is being assessed is what should be assessed. Validity therefore concerns both the instrument and assessment process and the challenge (cases) with which the candidate is tested. Ideally the content of the assessment should reflect the practitioner's own practice as closely as possible

Acceptability is the degree to which the assessment process is acceptable to all stakeholders. In tests of competence of a doctor the stakeholders are the doctor being assessed, the assessors, the people who provide the clinical challenge (patients or simulators), the profession, future patients of that doctor, and society

Feasibility is the degree to which the assessment can be delivered to all those who require it within real costs of staff and time constraints

Educational impact is the degree to which the assessment can assist the doctor to improve his or her performance, usually through the provision of feedback on specific strengths and weaknesses together with prioritised and specific strategies for improvement

from (Van der Vleuten, 1996; McKinley, Fraser *et al.*, 2001)

Appendix 8

Domains of assessment for quality in General Practice

RCGP searchable standards

This table lists the key areas of assessment of the work of a General Practitioner and his/her practice. The table lists the criteria within each area and indicates how compliance with the criterion is assessed in the different quality markers that the College offers. It has been compiled in this format by Mayur Lakhani based on a review of this topic undertaken by Alison Kay. The use of the word standard is in its broadest sense - namely the 'required level of quality'. The College has defined standards in most areas of the work of a GP and his/her practice. This table will be updated as and when new standards are published. e.g. out of hours.

Topics

1. [Access, availability and continuity of care](#)
2. [Privacy](#)
3. [Cervical cytology](#)
4. [Child health surveillance](#)
5. [Chronic disease management](#)
6. [Clinical audit](#)
7. [Communication with patients](#)
8. [Complaints management](#)
9. [Consultation length](#)
10. [Consultation skills](#)
11. [Contraceptive care](#)
12. [Emergency care](#)
13. [Help and information](#)
14. [Practice activity analysis](#)
15. [Practice library](#)
16. [Equipment](#)
17. [Health and safety](#)
18. [Records and Register](#)
19. [Referral letters](#)
20. [Terminal care](#)
21. [Health promotion](#)
22. [Home visiting](#)
23. [The future](#)
24. [Early diagnosis](#)
25. [Immunisation](#)
26. [Practice nurse](#)
27. [Pre-conceptual care](#)
28. [Professional values:](#)
29. [Premises and equipment](#)
30. [Referral rates](#)
31. [Repeat prescribing](#)
32. [Significant event review](#)
33. [Social dimension](#)
34. [Resources and use](#)
35. [Risk management office procedures](#)
36. [Team working](#)
37. [Use of locums](#)

Appendix 12

Further results of multilevel model

Category 1	Estimate	Standard Error
Fixed β	46.491	1.167
Random		
$\sigma^2_{\text{practice}}$	1.245	4.926
$\sigma^2_{\text{practice,doctor}}$	37.612	10.180
-2LogL	214.349	

Category 2	Estimate	Standard Error
Fixed β	26.308	0.890
Random		
$\sigma^2_{\text{practice}}$	0.981	2.867
$\sigma^2_{\text{practice,doctor}}$	20.377	5.522
-2LogL	194.524	

Category 3	Estimate	Standard Error
Fixed β	50.091	2.136
Random		
$\sigma^2_{\text{practice}}$	0.00	0.00
$\sigma^2_{\text{practice,doctor}}$	150.507	37.052
-2LogL	259.112	

Category 4	Estimate	Standard Error
Fixed β	17.697	0.671
Random		
$\sigma^2_{\text{practice}}$	0.00	0.00
$\sigma^2_{\text{practice,doctor}}$	14.878	3.663
-2LogL	182.746	

For the gender effect in all four categories:

Category 1	Estimate	Standard Error
Fixed		
β	49.254	1.449
Female	-6.556	1.844
Random		
$\sigma^2_{\text{practice}}$	4.903	5.794
σ^2_{female}	25.117	6.862
-2LogL	204.356	

Category 2	Estimate	Standard Error
Fixed		
β	23.348	0.971
Female	-5.297	1.330
Random		
$\sigma^2_{\text{practice}}$	1.664	2.491
σ^2_{female}	13.167	3.586
-2LogL	181.844	

Category 3	Estimate	Standard Error
Fixed		
β	55.633	2.432
Female	-14.398	3.609
Random		
$\sigma^2_{\text{practice}}$	6.375	14.776
σ^2_{female}	97.914	26.583
-2LogL	246.744	

Evaluating Process and Outcome in the Education of General Practitioners

Category 4	Estimate	Standard Error
Fixed		
β	18.714	0.789
Female	-2.798	1.308
Random		
$\sigma^2_{\text{practice}}$	0.00	0.00
σ^2_{female}	13.067	3.217
-2LogL	178.462	

Patient perception model runs

Category 1	Estimate	Standard Error
Fixed: β	57.722	0.742
Random:		
$\sigma^2_{\text{practice}}$	0.000	0.000
$\sigma^2_{\text{practice,doctor}}$	4.743	4.410
$\sigma^2_{\text{practice,doctor,patient}}$	99.262	9.381
-2LogL	1898.797	

Category 2	Estimate	Standard Error
Fixed: β	32.159	0.551
Random:		
$\sigma^2_{\text{practice}}$	0.000	0.000
$\sigma^2_{\text{practice,doctor}}$	2.172	2.434
$\sigma^2_{\text{practice,doctor,patient}}$	58.335	5.508
-2LogL	1761.823	

Category 3	Estimate	Standard Error
Fixed: β	39.361	0.956
Random:		
$\sigma^2_{\text{practice}}$	0.000	0.000
$\sigma^2_{\text{practice,doctor}}$	9.066	7.288
$\sigma^2_{\text{practice,doctor,patient}}$	152.290	14.456
-2LogL	1993.737	

Category 4	Estimate	Standard Error
Fixed: β	17.665	0.515
Random:		
$\sigma^2_{\text{practice}}$	0.000	0.000
$\sigma^2_{\text{practice,doctor}}$	0.000	0.000
$\sigma^2_{\text{practice,doctor,patient}}$	63.419	5.801
-2LogL	1670.047	

Including the factor “Importance” in each category of the patient-perceptions model gives the following results:

Category 1	Estimate	Standard Error
Fixed:		
β	33.573	4.113
ISUM 1	0.589	0.99
Random:		
$\sigma^2_{\text{practice}}$	0.509	1.550
$\sigma^2_{\text{practice,doctor}}$	0.000	0.000
$\sigma^2_{\text{practice,doctor,patient}}$	87.295	8.017
-2LogL	1776.907	

Category 2	Estimate	Standard Error
Fixed:		
β	9.268	2.443
ISUM 2	1.004	0.105
Random:		
$\sigma^2_{\text{practice}}$	0.293	0.779
$\sigma^2_{\text{practice,doctor}}$	0.000	0.000
$\sigma^2_{\text{practice,doctor,patient}}$	42.040	3.870
-2LogL	1592.993	

Category 3	Estimate	Standard Error
Fixed:		
β	7.125	3.252
ISUM 3	1.065	0.104
Random:		
$\sigma^2_{\text{practice}}$	0.594	2.892
$\sigma^2_{\text{practice,doctor}}$	7.270	6.050
$\sigma^2_{\text{practice,doctor,patient}}$	104.697	10.223
-2LogL	1811.318	

Evaluating Process and Outcome in the Education of General Practitioners

Category 4	Estimate	Standard Error
Fixed:		
β	2.281	1.657
ISUM 4	1.038	0.109
Random:		
$\sigma^2_{\text{practice}}$	0.000	0.000
$\sigma^2_{\text{practice,doctor}}$	0.616	1.706
$\sigma^2_{\text{practice,doctor,patient}}$	43.660	4.390
-2LogL	1497.851	

Further workings in MLM supporting findings in Chapter 5

We started with a simple variance component model, looking globally at whether **doctors rated themselves** lower or above the median (on a Performance scale of between 1 and 6, the median =3.5).

$$\begin{aligned} \text{meanpl}_{\text{doctor, practice}} &\sim N(XB, \Omega) \\ \text{meanpl}_{\text{doctor, practice}} &= \beta_{0\text{doctor, practice}} \text{cons} \\ \beta_{0\text{doctor, practice}} &= 0.389(0.093) + u_{0\text{practice}} + e_{0\text{doctor, practice}} \end{aligned}$$

$$[u_{0\text{practice}}] \sim N(0, \Omega_u) : \Omega_u = [0.005(0.031)]$$

$$[e_{0\text{doctor, practice}}] \sim N(0, \Omega_e) : \Omega_e = [0.253(0.068)]$$

$$-2 * \log\text{likelihood(IGLS)} = 48.955(33 \text{ of } 33 \text{ cases in use})$$

In ‘Patient Care’, doctors’ self-rating was 0.389 above the median of the performance scale (3.5). Ninety percent rated themselves as performing either at level 3 (moderately well) or 4 (well) on the original marking scale.

Whereas, obviously, differences in self-perception were found on the doctor level ($e_{0\text{doctors, practice}}$), the variation within practices is very similar to the overall population of 33 doctors. Differences approximating to zero between the practices, ($u_{0\text{practice}}=0.005$ SE 0.031), suggest that there might not be any differences on the practice level. However, a mean of zero might hide a lot of variation between the practices. To ensure we did not miss important variance, we calculated the contributions of each practice ($u_{0\text{practice}}$) to the population mean (0.389). This is shown below. The contributions are ranked. The contributions from left to right originate from practice numbers 3, 2, 7, 4, 6, 1 and 5, respectively. Ranking and displaying the error bars allows us to see whether the smallest contribution ($u_{0\text{practice}}$ at rank 1) is significantly different from its neighbour or any other practice.

In order to try to explain some of the differences, a new formula is adopted to work with the sum of the raw values in 'Patient Care'.

$$\begin{aligned}
 \text{psum1}_{\text{doctor, practice}} &\sim N(XB, \Omega) \\
 \text{psum1}_{\text{doctor, practice}} &= \beta_{0\text{doctor, practice}} \text{cons} \\
 \beta_{0\text{doctor, practice}} &= 46.591(1.167) + u_{0\text{practice}} + e_{0\text{doctor, practice}} \\
 [u_{0\text{practice}}] &\sim N(0, \Omega_u) : \Omega_u = [1.245(4.926)] \\
 [e_{0\text{doctor, practice}}] &\sim N(0, \Omega_e) : \Omega_e = [37.612(10.180)] \\
 -2*\loglikelihood(IGLS) &= 214.349(33 \text{ of } 33 \text{ cases in use})
 \end{aligned}$$

The results are similar to the first formula using mean values, with variance being found on the level of doctor but not of practice. Although less straightforward to interpret than the mean values of the previous model, the raw values are useful for comparing the effects of additional terms.

In the light of Dr Simonite's finding on the effect of gender on self-perception, we included the term female gender in the model, thus:

$$\begin{aligned}
 \text{psum1}_{\text{doctor, practice}} &\sim N(XB, \Omega) \\
 \text{psum1}_{\text{doctor, practice}} &= \beta_{0\text{doctor, practice}} \text{cons} + -6.556(1.844)\text{female}_{\text{doctor, practice}} \\
 \beta_{0\text{doctor, practice}} &= 49.254(1.449) + u_{0\text{practice}} + e_{0\text{doctor, practice}} \\
 [u_{0\text{practice}}] &\sim N(0, \Omega_u) : \Omega_u = [4.903(5.794)] \\
 [e_{0\text{doctor, practice}}] &\sim N(0, \Omega_e) : \Omega_e = [25.117(6.862)] \\
 -2*\loglikelihood(IGLS) &= 204.356(33 \text{ of } 33 \text{ cases in use})
 \end{aligned}$$

The female coefficient confirms that females rated themselves significantly lower than their male colleagues did. The deviance decreases significantly ($\chi^2 = 9.993$, 1dof, $p = 0.0015739$).

Does the tendency of male doctors to rate themselves higher vary between surgeries?

To answer this question, we need to introduce the random slope model, i.e. the slope is allowed to vary between surgeries. The model and the parameter estimates are shown below:

$$\text{psum1}_{ij} \sim N(XB, \Omega)$$

$$\text{psum1}_{ij} = \beta_{0ij}\text{cons} + \beta_{1ij}\text{female}_{ij}$$

$$\beta_{0ij} = 49.254(1.449) + u_{0ij} + e_{0ij}$$

$$\beta_{1ij} = -6.556(1.844) + u_{1ij}$$

$$\begin{bmatrix} u_{0ij} \\ u_{1ij} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 4.903(5.794) \\ 0.000(0.000) \quad 0.000(0.000) \end{bmatrix}$$

$$e_{0ij} \sim N(0, \Omega_e) : \Omega_e = [25.117(6.862)]$$

$$-2*\loglikelihood(IGLS) = 204.356(33 \text{ of } 33 \text{ cases in use})$$

Unfortunately, the model is working at the limits of its capacity in our small pilot study, with only 6 surgeries having more than 1 doctor, and we found inconsistent results on serial runs, meaning that we cannot make meaningful assertions on this point.

We looked to see if the model could be improved by including the factor Importance for category 1. This would show if doctors perceive themselves as performing differently for more or less important issues of the questionnaire.

$$\text{psum1}_{ij} \sim N(XB, \Omega)$$

$$\text{psum1}_{ij} = \beta_{0ij}\text{cons} + 0.229(0.266)\text{isum1}_{ij}$$

$$\beta_{0ij} = 37.728(10.319) + u_{0ij} + e_{0ij}$$

$$u_{0ij} \sim N(0, \Omega_u) : \Omega_u = [0.580(4.476)]$$

$$e_{0ij} \sim N(0, \Omega_e) : \Omega_e = [37.379(10.097)]$$

$$-2*\loglikelihood(IGLS) = 213.636(33 \text{ of } 33 \text{ cases in use})$$

Evaluating Process and Outcome in the Education of General Practitioners

The slope is not different to zero (0.229 with a S.E. of 0.266). Hence, including the Importance factor does not explain our model better. The independence of both variables is also shown graphically in figure 19.

In all four categories including a gender term improved the model significantly.

Table 5.11: Gender influence in the four categories

Category	χ^2 value (1 degree of freedom)	p-value
'Patient Care'	9.993	0.0015714
'Happy'	12.68	0.00036959
'Professional'	12.368	0.00043863
'Up-to-date'	4.284	0.038473

In 'Professional' ($\chi^2=5.6$, 1 degree of freedom (dof), $p=0.01796$) and 'Up-to-date' ($\chi^2=11.6$, 1 dof, $p=0.00065$) we found a significant influence of 'Importance'.

Running the model on the much more numerous patient perceptions of their doctors, we get following results for our basic model for 'Patient Care':

$$y_{ijk} \sim N(XB, \Omega)$$

$$y_{ijk} = \beta_{0ijk}x_0$$

$$\beta_{0ijk} = 57.722(0.743) + v_{0k} + u_{0jk} + e_{0ijk}$$

$$\begin{bmatrix} v_{0k} \end{bmatrix} \sim N(0, \Omega_v) : \Omega_v = \begin{bmatrix} 0.000(0.000) \end{bmatrix}$$

$$\begin{bmatrix} u_{0jk} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 4.743(4.410) \end{bmatrix}$$

$$\begin{bmatrix} e_{0ijk} \end{bmatrix} \sim N(0, \Omega_e) : \Omega_e = \begin{bmatrix} 99.262(9.381) \end{bmatrix}$$

$$-2*\loglikelihood(IGLS) = 1898.797(254 \text{ of } 254 \text{ cases in use})$$

In the fixed part we have a huge variation but, as was the case for self-perception, in the random part we cannot find a variation specifically for surgeries. The variation on level 2 is also similar to zero. Figure 20 illustrates this finding clearly.

Patients' perceptions of 'Patient Care' with inclusion of gender term:

$$psum1_{ijk} \sim N(XB, \Omega)$$

$$psum1_{ijk} = \beta_{0ijk} \text{cons} + -0.774(1.557) \text{gender}_{jk}$$

$$\beta_{0ijk} = 58.231(1.259) + v_{0k} + u_{0jk} + e_{0ijk}$$

$$[v_{0k}] \sim N(0, \Omega_v) : \Omega_v = [0.000(0.000)]$$

$$[u_{0jk}] \sim N(0, \Omega_u) : \Omega_u = [4.642(4.384)]$$

$$[e_{0ijk}] \sim N(0, \Omega_e) : \Omega_e = [99.236(9.377)]$$

$$-2 * \loglikelihood(IGLS) = 1898.550(254 \text{ of } 254 \text{ cases in use})$$

The deviance does not change substantially ($\chi^2=0.247$, $p=0.61920$).

Interestingly, this time, including the factor 'Importance' (ISUM1) does improve the model significantly.

$$psum1_{ijk} \sim N(XB, \Omega)$$

$$psum1_{ijk} = \beta_{0ijk} \text{cons} + 0.589(0.099) \text{isum1}_{ijk}$$

$$\beta_{0ijk} = 33.573(4.113) + v_{0k} + u_{0jk} + e_{0ijk}$$

$$[v_{0k}] \sim N(0, \Omega_v) : \Omega_v = [0.509(1.550)]$$

$$[u_{0jk}] \sim N(0, \Omega_u) : \Omega_u = [0.000(0.000)]$$

$$[e_{0ijk}] \sim N(0, \Omega_e) : \Omega_e = [87.295(8.017)]$$

$$-2 * \loglikelihood(IGLS) = 1776.907(243 \text{ of } 254 \text{ cases in use})$$

The deviance is significantly reduced ($\chi^2 = 121.89$, 1dof, $p=2.4400e-028$). Figure 24 shows that there is clearly some kind of correlation in patients' view about performance and importance and category one ('Patient Care')

Evaluating Process and Outcome in the Education of General Practitioners

Illegitimately modeling ‘Colleagues’ in the same way as we did patients, with three levels (colleague, doctor, practice) Dr Conradt obtained the following results for the basic model of ‘Patient Care’.

$$psum1_{colleague, doctor, practice} \sim N(XB, \Omega)$$

$$psum1_{colleague, doctor, practice} = \beta_{0colleague, doctor, practice} \text{cons}$$

$$\beta_{0colleague, doctor, practice} = 54.005(0.820) + v_{0practice} + u_{0doctor, practice} + e_{0colleague, doctor, practice}$$

$$[v_{0practice}] \sim N(0, \Omega_v) : \Omega_v = [0.000(0.000)]$$

$$[u_{0doctor, practice}] \sim N(0, \Omega_u) : \Omega_u = [4.821(5.141)]$$

$$[e_{0colleague, doctor, practice}] \sim N(0, \Omega_e) : \Omega_e = [173.090(13.641)]$$

$$-2 * \loglikelihood(IGLS) = 2805.403(350 \text{ of } 350 \text{ cases in use})$$

Including other variables, it looks like there are no differences in gender. There is a positive correlation with the variable “Importance” (p=5.6994e-005). This gives a hint of what might emerge, if we could analyze these data correctly.

Appendix 14

Semi-structured Interview Template (revised after pilot)

Recheck consent, ground-rules, in particular re-explore consent for interview to be taped and transcribed. Follow up all questions with requests for specific examples.

1. I'm interested in finding out specifically about the educational value to you of your training year in practice. How did the year in practice with your trainer, in your training practice, differ from just spending a year working as a GP in a non-training environment, gathering experience?
2. What if anything did you come away from your year feeling you'd learnt of particular relevance to later work?
3. What in particular, did you learn from working with your trainer?
4. Are there any ways that you think you've modelled yourself on him/her?
5. What about his/her teaching style...thinking about the positive aspects first of all, were there aspects that you found helpful about the way he/she taught?
6. Suggestions for improvement now, can you think of ways in which he/she could have made your learning experience more helpful? Are there ways in which he/she could have done it differently?
7. What about tutorials...are there any tutorials that stick in your mind as of lasting value?
8. Can you recall any unhelpful tutorials?
9. Thinking about the practice as a whole, were there any features of the practice, which were particularly helpful to you?
10. Can you think of any ways in which the practice could have prepared you better...were there aspects of the practice that could have been more helpful to you?
11. Looking back on the year do you think it was a year well spent really?
12. And would you have wanted more, less or was it about the right time really?
13. Any other comments, anything else I haven't covered about the training year you'd like to mention?

Appendix 15



Department of Primary Health Care

David Mant

Professor of General Practice, Head of Department

Ed Peile

Hon. Senior Clinical Lecturer (Medical Education)

Regina Conradt

Research Assistant

Institute of Health Sciences
Old Road, Headington

Oxford OX3 7LF

Telephone +44(0)1865 226750

Fax +44(0)1865 226621

Ed.peile@dphpc.ox.ac.uk

Regina.conradt@dphpc.ox.ac.uk

Version 3.1 Title of Project: Reflective work in General Practice – Towards a reliable tool to measure reflective thinking!

This questionnaire should take less than 5 minutes.

Dear Doctor,

Thank you for taking part in our research. This is an anonymous survey designed to identify typical factors influencing the self-perception of reflection in daily work. We have divided the questionnaire into two parts. First we will ask you a few questions about yourself and your background. Previous research has shown that self-perception of performance has a distinct gender bias. Whether other demographic data have an influence is yet unknown and we are looking for any unidentified factors. Second there are some statements about your learning behaviour during your daily work. Please rate each statement as honestly as possible. Remember, there are no right or wrong ratings, there are just different ways of dealing with daily work and challenges. If you have already filled in this questionnaire during another educational event, please **do not continue** and return the blank questionnaire.

Please **circle** the appropriate answer.

Recent post/status:

registrar principal locum physician in hospital other post please specify,.....

Age:

20-24 years 25-30 years 31-40 years 41-50 years 51-60 years >61 years

Gender:

male female

Ethnicity:

White (e.g. Caucasian, Hispanic) Asian or Pacific Islander Caribbean African Mixed Other, please specify

Do you take part in clinical education sessions on a regular basis?

Yes, at least one per months Yes, probably one each quarter Yes, probably one each year No, or very rarely.

Do you take part in non-clinical education sessions on a regular basis?

Yes, at least one per months Yes, probably one each quarter Yes, probably one each year No, or very rarely.

When did you graduate?

1-2 years ago 3-5 years ago 6-10 years ago 11-20 years ago > 21 years ago

Where did you graduate (country, town)?

Where do you work now (country, town)?

Instructions for answering the questionnaire

Each statement is followed by a series of possible ratings: definitely disagree, disagree with reservation, agree with reservation or definitely agree. Please rate each statement **by circling** the number that best represents your opinion about your activity (PERFORMANCE) in your daily work. Please, circle **Not Applicable** only if you really feel the statement does not apply to you and a definite answer is not possible.

Try to use the full range of rating scales. Make your strengths and weaker areas really stand out. Do not spend too long on each statement.

Performance

1 = definitely disagree, 2 = disagree (some reservation), 3 = agree (some reservation), 4 = definitely agree, N/A = Not Applicable

I need to understand all the pros and cons of a treatment before I'm happy to use it.	N/A	1	2	3	4
I like to have clear-cut guidelines.	N/A	1	2	3	4
I often reflect on my actions to see if I could have improved on what I did.	N/A	1	2	3	4
When something goes wrong, or I make a mistake, I try to understand why so as to be able to do better next time.	N/A	1	2	3	4
In the course of my practical work I have changed some of my beliefs.	N/A	1	2	3	4
I feel that the right way to do things is to follow established custom and well known practice.	N/A	1	2	3	4
I often re-appraise my experience so I can learn from it and improve on it next time.	N/A	1	2	3	4
I like to have an explanation of the way we do things.	N/A	1	2	3	4
I try to seek feedback about myself, even though this may be painful, because it can help me in achieving my goals.	N/A	1	2	3	4
When I face difficult situations in my present work I think back on the past on what has gone well and not so well and that is what influences how I proceed.	N/A	1	2	3	4
When something happens –pleasant or unpleasant- I try not only to experience it but also to learn from it.	N/A	1	2	3	4
I am not happy to base my practice of following the examples of others without having a deep understanding of the theoretical principles.	N/A	1	2	3	4
I need to understand the rationale for my actions.	N/A	1	2	3	4
If I follow what I have been taught I can deal well with most patients.	N/A	1	2	3	4
At my daily work in the surgery I do things so many times that I start doing them without thinking about it.	N/A	1	2	3	4
I often question the way others do something and try to think of a better way.	N/A	1	2	3	4
I try to remain open-minded and seek other people's views especially those that on the face of it seem to differ from, or contradict, my own.	N/A	1	2	3	4
When I work on common cases, I do so without thinking about what I am doing.	N/A	1	2	3	4
When making grave decisions I often try to put myself in the place of the other people who are involved, and I try to imagine how they will feel as a result of what I decide to do.	N/A	1	2	3	4
In my daily work I have to continually think about underlying concepts.	N/A	1	2	3	4

*Thank you very much for taking your time for filling in the questionnaire.
Your research team.*

Appendix 16

Self -Assessing your educational behaviours as a GP Trainer

Introduction

We have been researching assessment by the visiting team of trainers' educational behaviours (Peile *et al.*, 2000). Experience has shown that, whilst trainers welcome formatted feedback about the process by which they teach, the collation of evidence by the visiting team has proved difficult (Peile, 2003). The difficulty for visiting team was collecting and recording evidence across 8 categories of behaviours in the course of one short visit.

As the process of re-accreditation moves ever further along the path of reflective self-assessment that can be validated by the visiting team, it is highly desirable to extend this into the domain of 'The Trainer as Teacher'. It is hoped that trainers reflecting on their teaching methodically, will be able to find teaching behaviours that they would like to change, with consequent long-term benefit to the learner.

Rating your educating based on evidence

The way this process works is that you are given a framework within which to look at how you teach. The framework is made up of 8 categories of educational behaviours, with dimensions, which span between the sort of behaviours, which learners preferred in the long-term, and those behaviours, which they found less helpful in retrospect. You will gather evidence from looking at your own teaching in different ways, and you will record 'cues' to help remind you of the fragment of the evidence. Based on the amalgamated evidence, you will rate your overall behaviour, and your self-assessment will be available to the visitors at your re-accreditation. The visitors will share with you any insights they glean into your educational behaviours by considering the evidence they accumulate on the visit.

Collecting the evidence

It is important to emphasize that the areas to look for evidence are all ones which visitors have traditionally looked at on training assessment visits. The process by which visitors 'verify' your self-appraisal will not, therefore involve any change to the existing visit format.

A framework for evidence

The first step is to make sure you are familiar with the categories of training behaviours and their dimensions spanning what we call 'the preferred behaviour' to what we call 'the less helpful behaviour'. A summary is attached to this paper, and the full Medical Teacher paper will be downloadable from the PGMDE website at www.oxford-pgmde.co.uk If you are familiar with the eight categories it will become easy to record evidence on the work sheets.

Recording the evidence

What we suggest you do is just jot down a little *aide memoire* of any point which seems to be evidence. These 'cues' are just a phrase or quote, which will help you to remember a fragment of evidence. Make a quick decision as to where you see this evidence lying across the dimension – if it is more on the side of the preferred behaviour, jot it on the left of the relevant chart, or if it appears that this is a less helpful behaviour set it more on the right hand side. Don't agonise about this, just put it in where it seems right - you can change it later, when you are reviewing your evidence.

Collecting the evidence

It is important to emphasize that the areas to look for evidence are all ones which visitors have traditionally looked at on training assessment visits. The process by which visitors 'verify' your self-appraisal will not, therefore involve any change to the existing visit format.

To look for evidence, we suggest the following steps.

1. Log and Programme including audits, protocols and guidelines. Systematically go through the log and look at past, present and future programmes. What do theses tell you about the educational process? Think about the use that is made of audit, and how you use protocols and guidelines
2. Video of tutorial. Although you will only show maybe 8 to 10 minutes of tutorial to the visitors, try to record at least a couple of hours of tutorial(s) and review it all to look for evidence. You may find it helpful to debate some of this evidence with the visiting team, when they watch a short section of the video with you.
3. Registrar curriculum planning review On at least one occasion, (preferably more), when you are reviewing the log and curriculum planning with your registrar, use your 'second head' (Neighbour, 1999) to think about the evidence that emerges about the educational process. You can use your registrar to help with this.
4. As a trainer, reflect on your educational practice Finally, take some time out to reflect on the totality of your educational practice. Use any evidence that comes to mind, and try to synthesise what typifies your educational behaviours.

To help you in this process, we have mapped some of the types of evidence you may collect in the different areas – this list is not exhaustive.

Weighted average

Perceptions often involve a process of 'weighted averaging' (Anderson, 1981). You will need to spend a little bit of time reviewing the evidence you have collated. Having reviewed your evidence, you need to reach an overall judgement as to where to place your predominant behaviour on a scale of 0 to 10. We don't want to give you guidance on this, just use your own judgement on weighting the evidence that you've recorded.

Evaluating Process and Outcome in the Education of General Practitioners

How to use your self-assessment

At the end of this exercise, you will have a self-rating on a scale of 1 to 10 for each of the eight categories of training behaviours. You will share this with your visitors, who will attempt to objectively 'verify' your assessments on the small amount of evidence gleaned on your visit, and will hope to engage in helpful debate with you, as a means of developing your education.

More importantly, you have a means of looking at your own teaching. Are there any areas you want to change or develop? If so, you can map the changes over time by comparing your ratings.

You can also use ratings as a tool in your discussions in the trainer group – remember that it is very important that you model sensitive feedback when discussing training behaviours with other trainers! We hope it will open up your discussions into this highly relevant area of training.

My self –ratings

	Date	Date	Date	Date
Training or Education				
Style Spectrum				
Space for Reflection				
Modelling Personal Development and Team Skills				
Learning Cycles				
Family practice in context				
Control & Direction				
Feedback				

Suggestions for improvement

We hope you find it helpful to have a framework, and to bring attention to bear on your educational process, which is after all, a central component of teaching.

Neither framework nor procedure for self-assessment is yet perfect, and we welcome all suggestions for improvement. Please direct these to: ed.peile@dphpc.ox.ac.uk

Evaluating Process and Outcome in the Education of General Practitioners

1	Training or Education Ref: 1,2,3,4, (see below)	Problem-Based Approach Teaching based on approaches to problems which are not limited to present-day contexts (preferred behaviour)	Emphasis on Managing Disease Teaching focused on current policies for disease management (less helpful behaviour)
2	Style Spectrum Ref: 1,3,4 (see below)	Wide variety of styles Learner exposed to different consulting styles and role-models in tutorials (preferred)	Narrow range of styles Teaching dominated by personal style and behaviour of trainer (less helpful)
3	Space for Reflection Ref: 2,3,4, (see below)	Encouraging reflective practitioner Safe environment to learn from mistakes (preferred)	Protocol driven behaviour Black and white approach adopted where learner is expected to adhere to guidelines and elements of blame culture likely (less helpful)
4	Modelling Personal Development and Team Skills Ref: 1,2,3,4 (see below)	Personal development and team management skills taught Guided learning of skills like time management, assertiveness, boundary-setting (preferred)	No emphasis on team behaviours Little attempt is made to help learner understand the importance of team-working and the areas of personal development that are involved (less helpful)
5	Learning Cycles Ref: 1,3,4 (see below)	Learning cycles completed A culture exists in the practice where reflection, audit, assessment all promote change and re-evaluation (preferred)	Haphazard change Culture is reactive to external pressures, and little evidence of information about the practice inspiring meaningful change (less helpful)
6	Family practice in context Ref: 2,3,4 (see below)	Contextualised Learning Trainer introduces the broader dimensions of family and health expectations (preferred)	Emphasis on presenting problem Focus remains on sorting and shifting (less helpful)
7	Control & Direction Ref: 1,2,3,4 (see below)	Learner centred approach Trainer encourages progression towards self-directed learning (preferred)	Trainer centred approach Trainer fails to match style to learner's needs. Rigid structure, fixed diet (less helpful)
8	Feedback Ref: 1,2,3,4 (see below)	Sensitive feedback Both positive and negative feedback delivered where appropriate, stimulating confidence in the learner, and encouraging change (preferred)	Inappropriate criticism Feedback either inadequate or misplaced or poorly delivered, often not timely or specific enough to be useful to learner (less helpful)

Methods for obtaining evidence

1. Log and Programme (including audits, protocols and guidelines)
2. Video(s) of tutorial
3. Registrar = Curriculum planning review
4. Trainer = Your own reflection

Methods for obtaining evidence

1. Log and Programme

- emphasis on managing disease or problem-based approach?
- exposure to other doctors and other team members throughout the year?
- opportunities used for learning about self and teams?
- Time set aside for teaching audit and team skills?
- Programmed personal development teaching?
- evidence of movement towards self direction?
- Feedback recorded and tracked?

Audit

- Does practice/do I think/work/teach in complete audit cycles, with change management built-in to the process?
- Is practice using learner's project and previous learners' projects?

Practice protocols & guidelines

- reflective practice encouraged or 'straitjacket' approach?

1. Video(s) of tutorial

- emphasis on managing disease or problem-based approach?
- reflective practice encouraged?
- opportunities used for learning about self and teams?
- evidence of family contextualisation?
- ad-hoc evidence of trainer adapting to learner stage?
- sensitive feedback demonstrated?

1. Registrar curriculum planning review

- emphasis on managing disease or problem-based approach?
- exposure to other doctors and other team members throughout the year?
- reflective practice encouraged?
- opportunities used for learning about self and teams?
- learner aware of completed audit cycles in practice, and resultant change management?
- learner understands educational needs assessment?
- evidence of trainer encouraging learner direction within trainer control?
- comfortable about feedback and able to give examples of pos/neg?

1. Reflecting on your educational practice

- emphasis on managing disease or problem-based approach?
- reflective practice encouraged?
- opportunities used for learning about self and teams?
- Needs assessment a part of care process?
- Needs assessment a part of educational process?
- evidence of trainer encouraging learner direction within trainer control?
- comfortable about feedback and able to give examples of pos/neg?

Training or Education

1		Problem-Based Approach Teaching based on approaches to problems which are not limited to present-day contexts (preferred behaviour)		Emphasis on Managing Disease Teaching focused on current policies for disease management (less helpful behaviour)						
Methods Log Video Registrar Trainer		record your 'evidence cue' on left or right side to remind you how you viewed it								
Rating (circle one)	1	2	3	4	5	6	7	8	9	10

Style Spectrum

2	Wide variety of styles Learner exposed to different consulting styles and role-models in tutorials (preferred behaviour)		Teaching dominated by personal style and behaviour of trainer (less helpful behaviour)	Narrow range of styles						
Methods Log Registrar Trainer	record your 'evidence cue' on left or right side to remind you how you viewed it									
Rating (circle one)	1	2	3	4	5	6	7	8	9	10

Space for Reflection

3	Encouraging reflective practitioner Safe environment to learn from mistakes (preferred behaviour)		Protocol driven behaviour Black and white approach adopted where learner is expected to adhere to guidelines and elements of blame culture likely (less helpful behaviour)							
Methods Log Video Registrar Trainer Protocols Other	record your 'evidence cue' on left or right side to remind you how you viewed it									
Rating (circle one)	1	2	3	4	5	6	7	8	9	10

Modelling Personal Development and Team Skills

4	Personal development and team management skills taught Guided learning of skills like time management, assertiveness, boundary-setting (preferred behaviour)		No emphasis on team behaviours Little attempt is made to help learner understand the importance of team-working and the areas of personal development that are involved (less helpful behaviour)							
Methods Std scenario Log Video Trainer Registrar	record your 'evidence cue' on left or right side to remind you how you viewed it									
Rating (circle one)	1	2	3	4	5	6	7	8	9	10

Learning Cycles

5	Learning cycles completed A culture exists in the practice where reflection, audit, assessment all promote change and re-evaluation (preferred behaviour)		Haphazard change Culture is reactive to external pressures, and little evidence of information about the practice inspiring meaningful change (less helpful behaviour)							
Methods Log Registrar Trainer Audit	record your 'evidence cue' on left or right side to remind you how you viewed it									
Rating (circle one)	1	2	3	4	5	6	7	8	9	10

Family practice in context

6	Contextualised Learning Trainer introduces the broader dimensions of family and health expectations (preferred behaviour)			Emphasis on presenting problem Focus remains on sorting and shifting (less helpful behaviour)						
Methods Registrar Trainer Log Video	record your 'evidence cue' on left or right side to remind you how you viewed it									
Rating (circle one)	1	2	3	4	5	6	7	8	9	10

Control and Direction

7	Learner centred approach Trainer encourages progression towards self-directed learning (preferred behaviour)		Trainer centred approach Trainer fails to match style to learner's needs. Rigid structure -fixed diet (less helpful behaviour)							
Methods Registrar Video Trainer Log	record your 'evidence cue' on left or right side to remind you how you viewed it									
Rating (circle one)	1	2	3	4	5	6	7	8	9	10

Feedback

8	Sensitive feedback Both positive and negative feedback delivered where appropriate, stimulating confidence in the learner, and encouraging change (preferred behaviour)		Inappropriate criticism Feedback either inadequate or misplaced or poorly delivered, often not timely or specific enough to be useful to learner (less helpful behaviour)							
Methods Registrar Log Trainer Video	record your 'evidence cue' on left or right side to remind you how you viewed it									
Rating (circle one)	1	2	3	4	5	6	7	8	9	10